23. The method of claim 1, wherein the step of providing nucleotides comprises the step of providing at least one nucleotide selected from the group consisting of: inosine 5'-triphosphate, pyrrolopyrimidine ribonucleotide 5'-triphosphate, adenosine ribonucleotide 5'-triphosphate, thymidine ribonucleotide 5'-triphosphate, and modifications thereof.

A' cont'd

24. The method of claim 1, wherein the step of providing nucleotides comprises the step of providing at least one nucleotide selected from the group consisting of: 2-thiocytidine ribonucleotide 5'-triphosphate, guanosine ribonucleotide 5'-triphosphate, adenosine ribonucleotide 5'-triphosphate, thymidine ribonucleotide 5'-triphosphate, and modifications thereof.--

REMARKS

Claims 1-10 were examined in this case. Claims 1-10 stand rejected under 35 U.S.C. § 112, first paragraph, and claims 1-9 stand rejected under 35 U.S.C. § 112, second paragraph. The present Response amends claims 1, 3-8, and 10, cancels claim 9, and adds new claims 19-24. Each of the objections and rejections raised in the Office Action is addressed individually below.

Support for the Amendment

Support for the amendment to claim 1 can be found at page 4, lines 22-23 and page 15, lines 7-10. Support for the amendment to claims 3-8 can be found at page 4, lines 11-18. Support for the amendment to claim 10 can be found at page 7, lines 1-3. Support for new claims 19-24 can be found at page 18, lines 19-20 and page 25, lines 13-17. Finally, with respect to the claim language "and modifications thereof" added to claims 3-8 and 19-24, Applicant wishes to direct the Examiner's attention to page 9, line 11 to page 11 line 19, where what is meant by modifications of ribo- and deoxyribonucleotide bases is clearly defined. Based on these definitions, it will be appreciated that these modifications include modifications to the phosphate moiety, the ribose ring, or the nitrogenous base of the nucleotide. No new matter has been added by these amendments.

Rejections Under 35 U.S.C. § 112, Second Paragraph

Claims 1-9 stand rejected under 35 U.S.C. § 112, second paragraph, as being indefinite. This rejection has a number of aspects that are addressed in the order in which they appear below.

The Examiner asserts that the claims are drawn to methods of making "unstructured" nucleic acids, but the term "unstructured" is unclear since all nucleic acids have some structure. In response, Applicant points out that the term "unstructured nucleic acids" is defined in the specification. The specification states that

> "nucleic acid molecules having reduced levels of secondary structure compared to nucleic acid molecules of the same nucleotide sequence containing only naturally-occurring bases. . . are referred to herein as "unstructured nucleic acids" (UNAs). UNAs have reduced levels of secondary structure because of their reduced ability to form intramolecular hydrogen bond base pairs between regions of substantially complementary sequence. Preferred UNAs, however, retain the ability to form intermolecular hydrogen bond base pairs with other nucleic acid molecules" (page 15, lines 2-10).

The specification further describes what is meant by "unstructured nucleic acids" by stating that

> "UNAs contain nucleotide base analogs or a mixture of base analogs and naturally-occurring bases such that regions of sequence complementarity within the UNA are unable to form base pairs. One or both of the nucleotides that together form an intramolecular complementary base pair are substituted with a nucleotide containing a base analog so that the base pair is no longer formed, or is only formed at a reduced level. Preferably, the reduced level of base pairing is no more than one hydrogen bond interaction. Preferably, the analog(s) is selected so that the UNAs retain the ability to hybridize with another nucleic acid molecule of complementary or substantially complementary sequence" (page 15, lines 11-18).

As but one last example, Applicant points to page 22, lines 20-22, which state that "UNAs are produced such that sequence elements in the UNA have a reduced ability to hybridize to

substantially complementary sequence elements within the same UNA polynucleotide molecule." Indeed, the specification proceeds to describe the particular base pairing concepts of unstructured nucleic acids in detail at page 15, line 19 to page 17, line 3.

In light of the definition of unstructured nucleic acids provided in the specification, this aspect of the rejection should be withdrawn.

In another aspect of the rejection, the Examiner asserts that the claims further specify providing nucleotide precursors "sufficient" to synthesize an unstructured nucleic acid, which nucleic acid precursors include "pairs of complementary precursors that are unable to hybridize with one another." The Examiner questions whether both precursors are modified or whether just one precursor is modified.

In response, claim 1 has been amended to recite "'sufficient to synthesize' a nucleic acid strand complementary to at least a portion of the template nucleic acid strand." Applicant submits that the meaning of this phrase would be clear to one skilled in the art. One skilled in the art would recognize that this phrase means that nucleotide precursors of nucleic acid bases (such as adenine, thymine, guanine, and cytosine), which are complementary to at least a portion of the nucleic acid bases in the template strand, are added to a reaction mix so that they are incorporated into the growing nucleic acid strand by a polymerase. In addition, the phrase "pairs of complementary precursors" has been replaced to with language that sets forth that the nucleotide precursors include first and second complementary nucleotides that are complementary to each other, wherein the at least two complementary nucleotides have a reduced ability to form intramolecular base pairs but can form intermolecular base pairs.

Regarding the question of whether one or both of the nucleotides is modified, Applicant directs the Examiner's attention to page 4, lines 3-10, where it is set forth that either one or both of the nucleotides may be modified. Specifically, the specification teaches that

> "the present invention provides a method of producing nucleic acid molecules with reduced levels of intramolecular base pairing by incorporating *nucleotides* having modified bases such that complementary bases in a nucleic acid molecule are unable to form stable hydrogen bond base pairs. A modified base pair may

comprise *one* modified base and one natural base, or it may comprise *two* modified bases. Preferably, modified bases are positioned in nucleic acid molecules of the present invention in sequence elements of substantially complementary sequence to reduce intramolecular base pairing. Nucleic acid molecules of the present invention are produced by any method."

In light of these teachings, the claims appropriately contain no limitation of either one *or* two modified nucleotides, but encompass both possibilities and this aspect of the rejection may be withdrawn.

In light of the present amendment to the claims, and the teachings of the specification, Applicant asserts that this aspect of the rejection should be withdrawn.

In another aspect of the rejection, the Examiner asserts that the term "nucleotide precursors" is not defined such that one skilled in the art would understand the metes and bounds of the structures of the precursors. In addition, the Examiner points out that the phrase "nucleic acid precursors" in claim 1 has no antecedent basis. In response, although disagreeing with the Examiner's assertion that what is meant by the phrase "nucleotide precursors" would not be understood by one skilled in the art, Applicant has replaced the phrase "nucleotide precursors" with "nucleotides" for clarification. In addition, the phrase "nucleic acid precursors," which lacks antecedent basis, has been deleted. In light of these amendments, Applicant requests that this aspect of the rejection be withdrawn.

The Examiner further asserts that claims 2-8 refer to "nucleotides" and not "nucleotide precursors." In addition, the Examiner asserts that claims 3-8 specify particular precursors but also specify "or combinations thereof," which is unclear. In response, Applicant points out that the amendment to claim 1 replacing "nucleotide precursors" with "nucleotides" renders the first aspect of the rejection moot. In addition, with respect to use of the phrase "or combinations thereof," this phrase has been deleted from claims 3-8, and claims 3-8 have been amended to point out that at least one nucleotide is selected from the specified group. Applicant submits that one skilled in the art would know which nucleotide or combinations of nucleotides to select based on standard principles of Watson-Crick base pairing (see page 1, lines 5-13). In support of this assertion, Applicant points out that the specification teaches particular combinations of

nucleotides at page 4, lines 11-18 and more particularly page 15, line 19 to page 17, line 3. Some exemplary pairs of nucleotides are illustrated in Figure 1 of the drawing. In light of the above, withdrawal of this aspect of the rejection is requested.

In another aspect of the rejection, the Examiner asserts that, at the end of claim 1, the last step does not clearly relate back to the preamble of the claim. In particular, the Examiner asserts that the claim does not specifically claim synthesis of the "unstructured" nucleic acid. In response, claim 1 has been amended so that the end of claim 1 recites synthesis of an unstructured nucleic acid. Withdrawal of this aspect of the rejection is requested.

In another aspect of the rejection, the Examiner asserts that in claim 9, the preamble is not written correctly. Applicant apologizes for this oversight and submits that the preamble of claim 9 has been amended so that is it correctly written so that this aspect of the rejection may be withdrawn.

Finally, the Examiner states that claim 10 does not further limit claim 1 on the assertion that how the UNA is used does not further limit how it is made. Claim 10 has been rewritten as a method claim: a method of using an unstructured nucleic acid synthesized by the method of claim 1 by adding the unstructured nucleic acid to a ligase assay, a polymerase extension assay, or a nucleic acid array, each containing at least one nucleic acid template, and allowing the unstructured nucleic acid to hybridize to the nucleic acid template. In light of the above, withdrawal of this aspect of the rejection is requested.

Rejections Under 35 U.S.C. § 112, First Paragraph

Claims 1-10 stand rejected under 35 U.S.C. § 112, first paragraph, for lack of enablement. Specifically, the Examiner argues that the elected species of enzyme is an RNA polymerase, but the dCTP, dGTP, dATP and dTTP are the natural precursors for making a DNA nucleic acid, not an RNA nucleic acid, and dITP is often used in sequence reactions of DNA.

Claim 1 has been amended to recite a method of synthesizing an unstructured nucleic acid using an *RNA polymerase* and claim 9 has been canceled. In addition, claims 19-24 have been added to recite ribonucleotide triphosphates.

The method now claimed is fully enabled by the specification as filed for use of an RNA polymerase for synthesizing unstructured nucleic acids. First, the specification teaches that RNA polymerases may be used to generate RNA versions of unstructured nucleic acids. For example, the specification states that some "non-limiting examples of RNA polymerases suitable for generating RNA version of UNAs include the bacteriophage RNA polymerases from SP6, T7 and T3" (page 18, lines 19-20). In addition, the specification states that "although the base-pairing selectivity for these analog pairs has been experimentally tested for only DNA duplexes, it is likely that these same rules will hold for RNA duplexes and DNA/RNA heteroduplexes as well. This would allow for RNA versions of UNAs to be generated by transcription of PCR or cDNA products using the ribonucleotide triphosphate forms of the UNA analog pairs and RNA polymerases" (page 25, lines 13-17).

The Examiner asserts that an RNA polymerase will not incorporate deoxy precursors, but rather ribose precursors, such that one skilled in the art would necessarily practice an undue amount of experimentation to find modified precursors that would function to make the claimed unstructured nucleic acids. This is simply untrue. It was well known in the art at the time of the invention that nucleotides, other than ribonucleotides, could be incorporated into a nucleic acid template by an RNA polymerase at the time of the invention. For example, an abstract by Padilla ("Mutant T7 RNA-polymerase as a DNA-polymerase" *EMBO Journal* 14(18):4609-4621 September 15, (1995) (Exhibit A)), discloses the identification of a T7 RNA polymerase mutant that efficiently utilizes deoxyribonucleoside triphosphates. In addition, Sousa et al. ("Determinants of ribose specificity in RNA polymerization: Effects of Mn2+ and deoxynucleoside monophosphate incorporation into transcripts: *Biochemistry* 36(44):13718-13728 Nov. 4, (1997) (Exhibit B)), discloses that "mutation of tyrosine 639 [of T7 RNA polymerase] reduces specificity by a factor similar to 20 and largely eliminates the K-m differences between rNTPs and dNTPs." As but one last example, Sousa ("Efficient synthesis of nucleic acids heavily modified with non-cononical ribose 2' groups using a mutant T7 RNA polymerase (RNAP" 27(6):1561-1563 Mar 15 (1999) (Exhibit C)), discloses a T7 mutant (Y639F) that eliminates discrimination of the chemical character of the NTP ribose 2'-group to facilitate incorporation of non-canonical substrates into nucleic acids.

Based on these teachings, the skilled artisan at the time of the invention would recognize that any of a variety of deoxynucleotides, ribonucleotides, or modified deoxynucleotides or ribonucleotides could be used with any available RNA polymerase. Furthermore, as pointed out above, given that the specification includes teachings of polymerases other than SP6, T7 and T3,one skilled in the art at the time of the invention would have been aware that art available RNA polymerases, particularly mutant T7 RNA polymerases capable of incorporating deoxynucleotides, could be used to generate the variety of unstructured nucleic acids included in claims of the present scope.

In summary, the specification fully enables generation of RNA UNAs by applying the concepts taught for generating DNA UNAs, which would be clearly understood by the skilled artisan. Moreover, given the availability in the art of RNA polymerases capable of utilizing deoxynucleotides at the time of the invention, combined with the teachings of the specification, the skilled artisan would have understood that the specification enables claims of the present scope, particularly with respect to the use of an RNA polymerase for synthesizing an unstructured nucleic acid using any suitable modified precursor. In light of the above, withdrawal of this rejection is requested.

CONCLUSION

Applicant respectfully requests reconsideration of the claims, as amended herein. Please charge any fees that may be associated with this matter, or credit any overpayments, to our Deposit Account No. 50-1078.

Respectfully submitted,

Dated: October 16, 2001

Brenda Herschbach Jarrell, Ph.D.
Registered No. 39,223

Attorney for Applicants
CHOATE, HALL & STEWART
53 State Street
Boston, MA 02109
Telephone: 617-248-5000
Facsimile: 617-248-4000

VERSION WITH MARKINGS TO SHOW CHANGES MADE

In the Specification:

Paragraph beginning at line 15 of page 12 has been amended as follows:

Figure 2. **A)** DNA primer and template sequence used for the polymerase extension reaction. **B)** Phosphorimage of the 10% denaturing PAGE analysis of the polymerase extension reactions. The dNTP composition (A/D, T/S, G, C) and the polymerase present in each reaction are indicated. The positions of the ^{32}P-labeld primer (SEQ ID NO: 1) and 30-mer products (SEQ ID NO: 2) are indicated by arrows.

Paragraph beginning at line 21 of page 12 has been amended as follows:

Figure 3. **A)** The 6-mer DNA primer and template (SEQ ID NO: 3) sequence used to test the incorporation of the 2-amino-2'-deoxyadenosine triphosphate in a polymerase extension reaction. **B)** Phosphorimage of the 20% denaturing PAGE analysis of the polymerase extension reactions. The dATP and dDTP concentrations present in each reaction are indicated. The positions of the ^{32}P-labeled DNA 6-mer and 7-mer products are indicated by arrows. **C)** Graphic representation of the percentage of 6-mer DNA primer converted to 7-mer DNA product as a function of dNTP concentration.

Paragraph beginning at line 4 of page 13 has been amended as follows:

Figure 4. **A)** The 6-mer DNA primer and template (SEQ ID NO: 3) sequence used to test the incorporation of the 2-thiothymidine triphosphate in a polymerase extension reaction. **B)** Phosphorimage of the 20% PAGE analysis of the polymerase extension reactions. The dTTP and 2-thioTTP concentrations present in each reaction are indicated. The positions of the ^{32}P-labeled DNA 6-mer and 7-mer products are indicated by arrows. **C)** Graphic representation of the percentage of 6-mer DNA primer converted to 7-mer product as a function of dNTP concentration.

Paragraph beginning at line 11 of page 13 has been amended as follows:

Figure 5. **A)** The 6-mer DNA primer and template (SEQ ID NO: 3) sequences used to test the incorporation of the 2-amino-2'-deoxyadenosine and 2-thiothymidine triphosphate in the polymerase extension reaction. **B)** MALDI mass spectra of the polymerase extension reactions containing the indicated dNTP. Then m/z values for the 6-mer and 7-mer extension products are indicated. **C)** Table summarizing the predicted and measured m/z values for the 6-mer and 7-mer extension products.

Paragraph beginning at line 18 of page 13 has been amended as follows:

Figure 6. The scheme for generating single-stranded polynucleotides using a primer (SEQ ID NO: 4)/template(SEQ ID NO: 5)- dependent polymerase extension reaction (producing SEQ ID NO: 6) followed by digestion of the template DNA with λ exonuclease.

Paragraph beginning at line 1 of page 14 has been amended as follows:

Figure 8. Predicted secondary structures for three related 56-polynucleotide sequences containing either the four natural (A, G, C, T) nucleotides (SEQ ID NO: 7, SEQ ID NO: 9, and SEQ ID NOS: 11) or the 2-amino-2'-deoxyadenosine (D) and 2-thiothymidine (S) nucleotide substitutions (SEQ ID NO: 8, SEQ ID NO: 10, and SEQ ID NOS: 12).

Paragraph beginning at line 14 of page 14 has been amended as follows:

Figure 11. The DNA primer and template (SEQ ID NO: 13, SEQ ID NO: 14, SEQ ID NO: 6, and SEQ ID NO: 15) sequences used to test the effect of the polynucleotide secondary structure on the polymerase extension reaction. The arrows indicate the direction of the polymerase extension reaction. Sequences in bold in the first three templates are derived from the primer shown in Figure 6 for the polymerization reaction used to generate each single-stranded template.

In the Claims:

1. (Amended) A method of synthesizing an unstructured nucleic acid, the method comprising steps of:

providing a nucleic acid template strand including a first template sequence element and a second template sequence element that is substantially complementary to the first template sequence element;

providing a collection of nucleotide precursors sufficient to synthesize a nucleic acid strand complementary to at least a portion of the template nucleic acid strand, which portion includes the first and second template sequence elements, the collection including first and

second complementary nucleotides, wherein the first and second complementary nucleotides have a reduced ability to form an intramolecular base pair but can form an intermolecular base pair [an unstructured nucleic acid including a first complementary sequence element and a second complementary sequence element, complementary to the first and second sequence elements, respectively, which nucleic acid precursors include pairs of complementary precursors that are unable to hybridize with one another]; and

contacting the template and [precursors] the nucleotides with an RNA polymerase enzyme characterized by an ability to polymerize the [precursors] nucleotides under conditions and for a time sufficient for incorporation of the [precursor pairs] nucleotides to synthesize [into] the unstructured nucleic acid so that said first complementary sequence element and said second complementary sequence element of the unstructured nucleic acid do not interact with one another.

2. The method of claim 1, wherein the step of providing nucleotides comprises the step of providing at least one nucleotide having a purine analog and at least one nucleotide having a pyrimidine analog such that said purine analog and said pyrimidine analog are not capable of forming a stable hydrogen bonded base pair.

3. (Amended) The method of claim 1, wherein the step of providing nucleotides comprises the step of providing [nucleotides] at least one nucleotide selected from the group consisting of: 2-aminodeoxyadenosine 5'-triphosphate, 2-thiodeoxythymidine 5'-triphosphate, deoxyinosine 5'-triphosphate, deoxypyrrolopyrimidine 5'-triphosphate, 2-thiodeoxycytidine 5'-triphosphate, deoxyguanosine 5'-triphosphate, deoxycytidine 5'-triphosphate, deoxyadenosine 5'-triphosphate, deoxythymidine 5'-triphosphate[, and combinations thereof], and modifications thereof.

4. (Amended) The method of claim 1, wherein the step of providing nucleotides comprises the step of providing [nucleotides] at least one nucleotide selected from the group consisting of: 2-aminodeoxyadenosine 5'-triphosphate, 2-thiodeoxythymidine 5'-triphosphate, deoxyinosine 5'-triphosphate, deoxypyrrolopyrimidine 5'-triphosphate[, and combinations thereof], and modifications thereof.

5. (Amended) The method of claim 1, wherein the step of providing nucleotides comprises the step of providing [nucleotides] at least one nucleotide selected from the group consisting of: 2-aminodeoxyadenosine 5'-triphosphate, 2-thiodeoxythymidine 5'-triphosphate, deoxyguanosine 5'-triphosphate, 2-thiodeoxycytidine 5'-triphosphate[, and combinations thereof], and modifications thereof.

6. (Amended) The method of claim 1, wherein the step of providing nucleotides comprises the step of providing [nucleotides] at least one nucleotide selected from the group consisting of: 2-aminodeoxyadenosine 5'-triphosphate, 2-thiodeoxythymidine 5'-triphosphate, deoxyguanosine 5'-triphosphate, deoxycytidine 5'-triphosphate[, and combinations thereof], and modifications thereof.

7. (Amended) The method of claim 1, wherein the step of providing nucleotides comprises the step of providing [nucleotides] at least one nucleotide selected from the group consisting of: deoxyinosine 5'-triphosphate, deoxypyrrolopyrimidine 5'-triphosphate, deoxyadenosine 5'-triphosphate, deoxythymidine 5'-triphosphate[, and combinations thereof], and modifications thereof.

8. (Amended) The method of claim 1, wherein the step of providing nucleotides comprises the step of providing [nucleotides] at least one nucleotide selected from the group consisting of: 2-thiodeoxycytidine 5'-triphosphate, deoxyguanosine 5'-triphosphate, deoxyadenosine 5'-triphosphate, deoxythymidine 5'-triphosphate[, and combinations thereof], and modifications thereof.

9. (Canceled) The method of one of claims 1-8, wherein the step of contacting with an enzyme comprises with an enzyme selected from the group consisting of: an RNA polymerase, a DNA polymerase, a reverse transcriptase, a ribozyme, and a self-replicating RNA molecule.

10. (Amended) [The] A method of using an unstructured nucleic acid synthesized by the method of claim 1, [wherein the unstructured nucleic acid is used in] comprising the steps of:

adding the unstructured nucleic acid to a ligase assay, a polymerase extension assay, or a nucleic acid array containing at least one nucleic acid template, and allowing the unstructured nucleic acid to hybridize to the nucleic acid.

Add the following new claims.

--19. The method of claim 1, wherein the step of providing nucleotides comprises the step of providing at least one nucleotide selected from the group consisting of: 2-aminoadenosine ribonucleotide 5'-triphosphate, 2-thiothymidine ribonucleotide 5'-triphosphate, inosine 5'-triphosphate, pyrrolopyrimidine ribonucleotide 5'-triphosphate, 2-thiocytidine ribonucleotide 5'-triphosphate, guanosine ribonucleotide 5'-triphosphate, cytidine ribonucleotide 5'-triphosphate,

adenosine ribonucleotide 5'-triphosphate, deoxythymidine 5'-triphosphate, and modifications thereof.

20. The method of claim 1, wherein the step of providing nucleotides comprises the step of providing at least one nucleotide selected from the group consisting of: 2-aminoadenosine ribonucleotide 5'-triphosphate, 2-thiothymidine ribonucleotide 5'-triphosphate, inosine 5'-triphosphate, pyrrolopyrimidine ribonucleotide 5'-triphosphate, and modifications thereof.

21. The method of claim 1, wherein the step of providing nucleotides comprises the step of providing at least one nucleotide selected from the group consisting of: 2-aminoadenosine ribonucleotide 5'-triphosphate, 2-thiothymidine ribonucleotide 5'-triphosphate, guanosine ribonucleotide 5'-triphosphate, 2-thiocytidine ribonucleotide 5'-triphosphate, and modifications thereof.

22. The method of claim 1, wherein the step of providing nucleotides comprises the step of providing at least one nucleotide selected from the group consisting of: 2-aminoadenosine ribonucleotide 5'-triphosphate, 2-thiothymidine ribonucleotide 5'-triphosphate, guanosine ribonucleotide 5'-triphosphate, cytidine ribonucleotide 5'-triphosphate, and modifications thereof.

23. The method of claim 1, wherein the step of providing nucleotides comprises the step of providing at least one nucleotide selected from the group consisting of: inosine 5'-triphosphate, pyrrolopyrimidine ribonucleotide 5'-triphosphate, adenosine ribonucleotide 5'-triphosphate, thymidine ribonucleotide 5'-triphosphate, and modifications thereof.

24. The method of claim 1, wherein the step of providing nucleotides comprises the step of providing at least one nucleotide selected from the group consisting of: 2-thiocytidine ribonucleotide 5'-triphosphate, guanosine ribonucleotide 5'-triphosphate, adenosine ribonucleotide 5'-triphosphate, thymidine ribonucleotide 5'-triphosphate, and modifications thereof.—

Exhibit A

MUTANT T7 RNA-POLYMERASE AS A DNA-POLYMERASE

SOUSA R, PADILLA R

EMBO JOURNAL
14 (18): 4609-4621 SEP 15 1995

Document type: Article **Language:** English **Cited References:** 35 **Times Cited:** 70

Abstract:
We have identified a T7 RNA polymerase (RNAP) mutant that efficiently utilizes deoxyribonucleoside triphosphates. In vitro this mutant will synthesize RNA, DNA or 'transcripts' of mixed dNMP/rNMP composition depending on the mix of NTPs present in the synthesis reaction. The mutation is conservative, changes Tyr639 within the active site to phenylalanine and does not affect promoter specificity or overall activity. Non-conservative mutations of this tyrosine also reduce discrimination between deoxyribo- and ribonucleoside triphosphates, but these mutations also cause large activity reductions. Of 26 mutations of other residues in and around the active site examined none showed marked effects on rNTP/dNTP discrimination. Mutations of the corresponding tyrosine in DNA polymerase (DNAP) I increase miscoding, though effects on dNTP/rNTP discrimination for the DNAP I mutations have not been reported. This conserved tyrosine may therefore play a similar role in many polymerases by sensing incorrect geometry in the structure of the substrate/template/product due to inappropriate substrate structure or mismatches. T7 RNAP can use RNA templates as well as DNA templates and is capable of both primer extension and de novo initiation. The Y639F mutant retains the ability to use RNA or DNA templates. Thus this mutant can display de novo initiated or primed DNA-directed DNA polymerase, reverse transcriptase, RNA-directed RNA polymerase or DNA-directed RNA polymerase activities depending simply on the templates and substrates presented to it in the synthesis reaction.

Author Keywords:
MUTAGENESIS, RNA POLYMERASE, SUBSTRATE SPECIFICITY, TRANSCRIPTION

KeyWords Plus:
REVERSE-TRANSCRIPTASE, CRYSTAL-STRUCTURE, ANGSTROM RESOLUTION, T7-RNA POLYMERASE, ESCHERICHIA-COLI, KLENOW FRAGMENT, I KLENOW, ELONGATION, SITE

Addresses:
SOUSA R, UNIV TEXAS,HLTH SCI CTR,DEPT BIOCHEM,7702 FLOYD CURL DR,SAN ANTONIO,TX 78212

Publisher:
OXFORD UNIV PRESS UNITED KINGDOM, OXFORD

IDS Number:
RX441

Exhibit B

HOME | HELP | DATE & DB LIMITS | GENERAL SEARCH | CITED REF SEARCH | MARK | LOG OFF

General Search Results--Full Record

Article 31 of 77 | PREVIOUS | NEXT | SUMMARY | FIND RELATED RECORDS | Explanation

Determinants of ribose specificity in RNA polymerization: Effects of Mn2+ and deoxynucleoside monophosphate incorporation into transcripts

Huang Y, Beaudry A, McSwiggen J, **Sousa R**

BIOCHEMISTRY
36 (44): 13718-13728 NOV 4 1997

Document type: Article **Language:** English **Cited References:** 56 **Times Cited:** 11

Abstract:
The catalytic specificity of T7 RNA polymerase (RNAP) for ribonucleoside triphosphates vs deoxynucleoside triphosphates {(k(cat)/K-m)(rNTP)/(k(cat)/K-m)(dNTP)} during transcript elongation is similar to 80. Mutation of tyrosine 639 to phenylalanine reduces specificity by a factor of similar to 20 and largely eliminates the K-m difference between rNTPs and dNTPs. The remaining specificity factor of similar to 4 is k(cat)-mediated and is nearly eliminated if Mn2+ is substituted for Mg2+ in the reaction. Mn2+ substitution does not significantly affect the K-m difference between rNTPs and dNTPs. Mn2+ substitution also enhances the activity of poorly active mutant enzymes carrying nonconservative substitutions in the active site, and its effects are generally consistent with the Mn2+–catalyzed reaction being less restrictive in its requirements for alignment of the reactive groups. In addition to discrimination occurring at the level of nucleoside monophosphate (NMP) incorporation, it is also found that transcripts containing deoxynucleoside monophosphates (dNMPs) are more poorly extended than transcripts of canonical structure, though a severe barrier to transcript extension is seen only when the 3' region of the transcript is heavily substituted with dNMPs. The barrier to extension of transcripts heavily substituted with dNMPs is reduced for sequences known to be amenable to forming A-like helices and is larger for sequences that resist transformation from B-form DNA . DNA structures. The barrier to extension of dNMP-substituted transcripts is also reduced by solution conditions known to destabilize B-form DNA and to stabilize A-form structures. These observations imply a requirement for a non-B-form, possibly A-like, conformation in the transcript template hybrid that is disrupted when the transcript is of predominantly deoxyribose structure.

KeyWords Plus:
DNA-POLYMERASE-I, ESCHERICHIA-COLI, REVERSE-TRANSCRIPTASE, ACTIVE-SITE, MANGANESE MUTAGENESIS, MOLECULAR-STRUCTURE, SYNTHETIC DNA, RESIDUES, FRAGMENT, MUTANTS

Addresses:
UNIV TEXAS,HLTH SCI CTR,DEPT BIOCHEM,SAN ANTONIO,TX 78284
RIBOZYME PHARMACEUT INC,BOULDER,CO 80301

Publisher:
AMER CHEMICAL SOC, WASHINGTON

IDS Number:
YE774

Exhibit C

General Search Results--Full Record

Efficient synthesis of nucleic acids heavily modified with non-canonical ribose 2 '-groups using a mutant T7 RNA polymerase (RNAP)

Padilla R, **Sousa R**

NUCLEIC ACIDS RESEARCH
27 (6): 1561-1563 MAR 15 1999

Document type: Article **Language:** English **Cited References:** 13 **Times Cited:** 4

Abstract:
A T7 RNAP mutant (Y639F) which eliminates discrimination of the chemical character of the NTP ribose 2'-group, facilitates incorporation of non-canonical substrates into nucleic acids. However, transcripts containing a high percentage of non-canonical NMPs are poorly extended due to effects of the 2'-substituents on the transcript:template hybrid conformation. We tested the addition of compounds that stabilize A-type helix geometry to the reaction. High concentrations of polyamines, together with other changes in reaction conditions, greatly increased the synthesis of transcripts heavily substituted with non-canonical ribose 2'-groups. Template structures that facilitate promoter opening increased the efficiency of reactions where non-canonical substrates were incorporated during transcription of +1 to +6.

KeyWords Plus:
DNA-POLYMERASE, CONFORMATION

Addresses:
Sousa R, Univ Texas, Hlth Sci Ctr, Dept Biochem, 7703 Floyd Curl Dr, San Antonio, TX 78284 USA
Univ Texas, Hlth Sci Ctr, Dept Biochem, San Antonio, TX 78284 USA

Publisher:
OXFORD UNIV PRESS, OXFORD

IDS Number:
178GA

ISSN:
0305-1048

3904–3909 Nucleic Acids Research, 2000, Vol. 28, No. 20

Exhibit A

Minimising the secondary structure of DNA targets by incorporation of a modified deoxynucleoside: implications for nucleic acid analysis by hybridisation

Hong-Khanh Nguyen and Edwin M. Southern*

Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU, UK

Received June 28, 2000; Revised and Accepted August 21, 2000

ABSTRACT

Some regions of nucleic acid targets are not accessible to heteroduplex formation with complementary oligonucleotide probes because they are involved in secondary structure through intramolecular Watson–Crick pairing. The secondary conformation of the target may be destabilised to assist its interaction with oligonucleotide probes. To achieve this, we modified a DNA target, which has self-complementary sequence able to form a hairpin loop, by replacing dC with *N*4-ethyldeoxycytidine (d^{4Et}C), which hybridises specifically with natural dG to give a G:^{4Et}C base pair with reduced stability compared to the natural G:C base pair. Substitution by d^{4Et}C greatly reduced formation of the target secondary structure. The lower level of secondary structure allowed hybridisation with complementary probes made with natural bases. We confirmed that hybridisation could be further enhanced by modifying the probes with intercalating groups which stabilise the duplex.

INTRODUCTION

Arrays of oligonucleotides are used for the analysis of gene expression (1–3), resequencing genes to find mutations and for polymorphism detection (4). However, a number of problems must be addressed before they can be used optimally. One limitation of the method is that stable secondary structures in the target nucleic acid can make the target sequence inaccessible to intermolecular Watson–Crick base pairing (5–7). Several techniques have been explored to alleviate this problem, but none appears to be totally successful. Fragmenting the nucleic acid sequence, preferably to a size close to that of the oligonucleotides on the array, by heating the DNA or RNA in the presence of Mg^{2+} (8) or by the use of uracil-*N*-glycosylase (9) can reduce the secondary structure effect. However, the extent of fragmentation is difficult to control. Oligonucleotide analogues such as peptide nucleic acid (PNA) exhibit PNA–DNA heteroduplexes with a high thermal stability at low salt concentration. The authors suggest that under these conditions the DNA strands are less likely to fold to form stable secondary structure and therefore would be accessible to the oligonucleotide probes (10). However, there is an increasing tendency for non-specific DNA binding. We here describe a new strategy for minimising the secondary structure in the target DNA by introduction of modified nucleosides into the target that form weak base pairs and show that targets modified in this way have a good affinity for oligonucleotide probes.

MATERIALS AND METHODS

Phosphoramidite monomers were from Cruachem. *N*4-ethyl-2′-deoxycytidine (d^{4Et}C) and 6-chloro-2-methoxyacridine phosphoramidites were from Glen Research. T4 polynucleotide kinase was from BioLabs. [γ-^{32}P]ATP was from Amersham. Acrylamide solution (40%, 19:1 acrylamide:bisacrylamide) was from Anachem (Luton, UK).

Oligonucleotides preparations

Chain assembly was carried out on an Applied Biosystems 394 synthesiser on a CPG (controlled pore glass) solid support functionalised with a nucleoside using phosphoramidite chemistry. Oligonucleotides **1–3** were deprotected by overnight treatment with concentrated ammonia solution at 55°C. Oligonucleotide **4** was deblocked using a deprotection procedure with 0.4 M methanolic sodium hydroxide (4:1 methanol/water) for 16 h at room temperature to avoid cleavage of the bond between the C9 atom of the acridine ring and the N atom of the linker (11). The oligonucleotides were then purified on a 15% denaturing polyacrylamide gel (TBE, 150 V, 5 h). The oligonucleotides (10 pmol) were 5′-terminal-labelled with T4 polynucleotide kinase (5 U) for 1 h at 37°C in the presence of [γ-^{32}P]ATP.

Melting temperature experiment

Changes in absorbance with temperature of 2 µM oligonucleotides in 10^{-2} M sodium cacodylate buffer, pH 7, containing 1 M NaCl and 2×10^{-4} M EDTA were measured at λ = 260 nm on a Beckman DU 640 cell changer spectrophotometer. The samples and the reference were slowly heated at a rate of 0.5°C/min from 15 to 90°C. Melting temperature (T_m) was taken as the temperature corresponding to half-dissociation of the structure.

*To whom correspondence should be addressed. Tel: +44 1865 275282; Fax: +44 1865 275283; Email: ems@bioch.ox.ac.uk
Present address:
Hong-Khanh Nguyen, Proteus, Parc Georges Besse, Allée Graham Bell, 30 000 Nimes, France

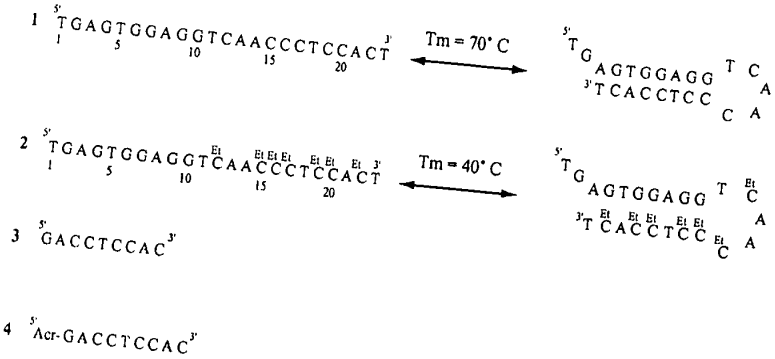


Figure 1. Single-stranded structure in equilibrium with the hairpin structure of unmodified 23mer oligodeoxynucleotide **1** and modified 23mer oligodeoxynucleotide **2** involving $d^{4Et}C$. T_m was determined at an oligomer strand concentration of 2 μM in 10^{-2} M sodium cacodylate buffer, pH 7, containing 1 M NaCl and 2×10^{-4} M EDTA. Sequences of the nonamer oligodeoxynucleotide probe **3** and probe **4** bearing at its 5′-end a 6-chloro-2-methoxyacridine molecule (Acr). The acridine molecule is covalently attached to the 5′-terminus of the oligonucleotide via a hexaethylene spacer. All oligodeoxynucleotide sequences are in the 'deoxy' series.

Electrophoresis

Aliquots of 0.5 pmol oligonucleotides **1** and **2** were mixed with the appropriate amount of probes **3** and **4** in 10^{-2} M sodium cacodylate buffer, pH 7, containing 20 mM NaCl, 30% glycerol and bromophenol blue. Samples were heated to 90°C and then cooled slowly before being kept at 4°C for 30 min prior to loading. PAGE experiments were carried out for 3 h on a 15% non-denaturing gel in TBE buffer, pH 7, containing 20 mM NaCl at 4°C. Gels were exposed to a storage phosphor plate and scanned using a phosphorimager (Molecular Dynamics Storm 860).

Fabrication of arrays

The oligonucleotide array was prepared directly on the surface of aminated polypropylene (12) in an adapted Applied Biosystem's DNA synthesiser (ABI) using a 49.5 mm diamond-shaped mask as described previously (13–14). Standard phosphoramidites were used in synthesis. The offset between each of the 23 coupling steps was 1.98 mm. The arrays were deprotected in 30% ammonia solution at 55°C for 16 h in a closed chamber.

Hybridisation conditions and analysis

Hybridisations were performed in 1 M NaCl in the presence of 10 mM Tris, pH 7, 1 mM EDTA and 0.01% SDS at 4°C for 3 h. Approximately 2 pmol of oligonucleotide was diluted in 4 ml hybridisation solution. The array was then briefly washed in the hybridisation solution and dried with tissue paper. Autoradiography and analysis were carried out as described (15). The array was stripped in 100 mM sodium carbonate/bicarbonate buffer (pH 10) containing 0.01% SDS for 2–3 min at 90–95°C and re-used.

RESULTS

Our strategy involves modification of the target, incorporating nucleic acid analogues to reduce the stability of intramolecular base pairs. For practical applications it will be necessary to incorporate the modified bases in the nucleic acid fragments to be analysed by an enzymatic reaction, constraining the choice of modified nucleosides to those whose 5′-triphosphates can be used as substrates for DNA polymerases. From the large number of modified nucleosides described in the literature we chose $d^{4Et}C$ to replace natural deoxycytidine (dC); $d^{4Et}C$ is known to hybridise specifically with natural deoxyguanosine (dG) to give a G:^{4Et}C base pair having reduced stability as compared to the natural G:C base pair, but very close to that of A:T. Moreover, it has been shown that its triphosphate can be used as a substrate by DNA polymerases (16).

Choice of target sequence

Most secondary structures include hairpin loops composed of duplex stems and a terminal loop of unpaired nucleotides (17–18). Because GC-rich sequences are more likely to form stable intramolecular structures, we chose to decrease the stability of the G:C base pair. The target sequence is 23 nt in length; an unmodified sequence containing only natural nucleosides (sequence **1**) and a modified sequence including $d^{4Et}C$ instead of dC (sequence **2**). Both sequences can be folded to form a hairpin structure containing a stem of 8 bp (five of which are G:C or G:^{4Et}C base pairs) and a terminal loop of 5 nt (Fig. 1).

Effect of $d^{4Et}C$ on the stability of the DNA target secondary structure

The stability of the hairpin structure was determined by absorption spectroscopy. In 1 M NaCl the native sequence **1** exhibited a half-dissociation temperature (T_m) of 70°C, clearly higher than the T_m of the modified oligonucleotide **2**, which was 40°C.

The structures of the native and $d^{4Et}C$-modified oligonucleotides **1** and **2** were analysed by mobility shift on a non-denaturing polyacrylamide gel (Fig. 2). The unmodified oligonucleotide **1** (lane 1) migrated as a single band faster than the linear oligonucleotide reference 23mer oligodeoxythymidylate (T_{23}) (lane 5), suggesting that it contained a hairpin. The $d^{4Et}C$-modified oligonucleotide **2** also migrated as a major band at the same rate as unmodified oligonucleotide **1** but also showed

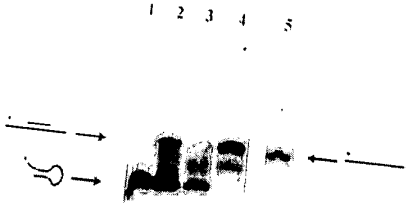


Figure 2. Electrophoretic analysis of ^{32}P-labelled sequences **1** and **2** (0.5 pmol) in the absence and presence of nonadeoxynucleotide probes **3** and **4** complementary to the region G4–C13 (^{4Et}C13). PAGE experiments were carried out for 3 h on a 15% non-denaturing gel in TBE buffer, pH 7, and containing 20 mM NaCl at 4°C. Lane 1, sequence **1**; lane 2, sequence **1** + 600 equiv. probe **4**; lane 3, sequence **2**; lane 4, sequence **2** + 600 equiv. probe **4**; lane 5, T_{23}.

a smear between the positions of the hairpin and linear molecule, suggesting an equilibrium between the two forms (lane 3).

To see if a short probe can displace the target secondary structure to form a heteroduplex, mixtures of the targets with a large excess of nonadeoxynucleotide probe bearing a 6-chloro-2-methoxyacridine molecule at its 5′-end (oligonucleotide **4**) were analysed by polyacrylamide gel retardation (Fig. 2). We chose to add the acridine moiety because a short oligonucleotide might not have a strong enough affinity for its target sequence and it has been shown that affinity can be greatly increased by covalently linking an intercalating molecule, like acridine, at the end of the oligonucleotide (19–20). In the presence of probe **4** the unmodified sequence **1** presented two well-resolved bands (Fig. 2, lane 2). The fast moving band corresponding to the hairpin structure was much more intense than the slow moving band attributed to the heteroduplex formed with the acridine-modified probe **4**. In contrast, the d^{4Et}C-modified target (sequence **2**) in the presence of the acridine probe **4** migrated as a unique retarded band corresponding to the heteroduplex (lane 4). Under these experimental conditions the unmodified target **1** exists primarily as a hairpin structure and a large excess of probe **4** (600 equiv.) did not completely displace the intramolecular structure to heteroduplex formation. We estimate that only 30% of unmodified sequence **1** was involved in heteroduplex, in contrast to the d^{4Et}C-modified sequence **2** containing d^{4Et}C; for this target only the linear duplex was formed.

The quantity of probe necessary to induce complete hetero-duplex formation was estimated by gel retardation analysis of d^{4Et}C-modified oligonucleotide **2** in the presence of varying concentrations of the acridine probe **4**. It was found that 1 equiv. of probe **4** is sufficient to open most of the intra-molecular structure, allowing the formation of heteroduplex. The experiment was also carried out with a nonadeoxynucleotide probe lacking the acridine moiety (sequence **3**). In this case ~10 equiv. of probe **3** were necessary to form the linear duplex, confirming the stabilising influence of the acridine. We noted that this mixture produced a smear, suggesting that the hetero-duplex might not be very stable under the experimental conditions (data not shown).

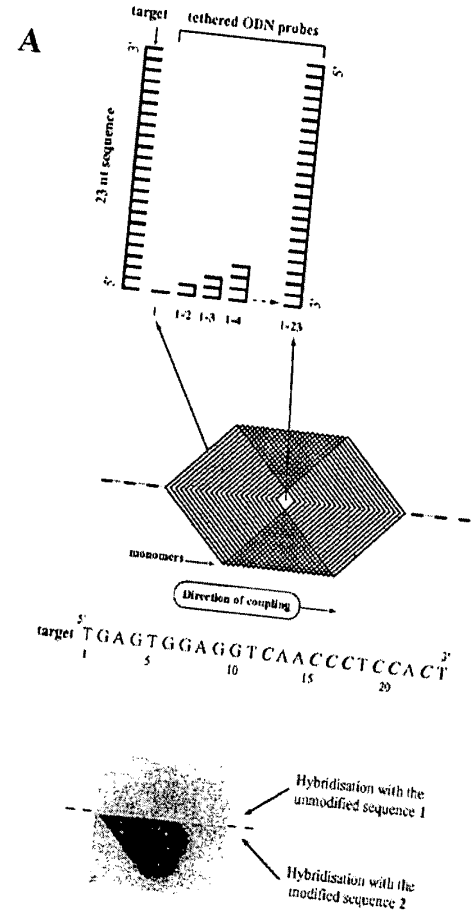


Figure 3. Organisation of a scanning array. (**A**) A template of overlapping diamonds representing the footprints of a diamond-shaped mask used to make a scanning array. The tiling path on the top shows the positions of some of the oligonucleotides in the grid. Successive chevrons are occupied by oligonucleotides differing in length by 1 nt. The series of oligonucleotides is made by coupling the bases through a mask to a polypropylene surface in the order in which they occur in the complement of the target sequence, starting from the first base at the 5′-end. Synthesis is carried out on an automated DNA synthesiser, such as an ABI 394. Using standard nucleotide phosphoramidites in the synthesis, the oligonucleotides are attached to the solid support at their 3′-ends. After each coupling step the mask is moved along the substrate by a 1.98 mm size step. The process is continued until an array of 23mer length is obtained. (**B**) Image obtained after hybridisation of ^{32}P-labelled unmodified sequence **1** (top half) and d^{4Et}C-modified sequence **2** (bottom half) to the complementary oligonucleotide scanning array. Hybridisations were performed in 1 M NaCl in the presence of 10 mM Tris, pH 7, 1 mM EDTA and 0.01% SDS at 4°C for 3 h. Approximately 2 pmol of oligonucleotide was diluted in 4 ml of hybridisation solution. The array was then washed briefly in the hybridisation solution, dried with tissue paper and exposed to a phosphor storage screen and scanned with a phosphorimager. *C* = C or ^{4Et}C.

Hybridisation of the unmodified and modified targets to a scanning array

To further characterise the effect of d^{4Et}C on the stability of secondary structure and the effect on its ability to form

heteroduplex, unmodified oligonucleotide **1** and $d^{4Et}C$-modified oligonucleotide **2** were hybridised to a scanning array of complementary oligonucleotides made with natural nucleotides (13–15). The scanning array of oligonucleotides, ranging in size from monomers to 23mers, represented complements of both 23 nt targets (sequences **1** and **2**). The organisation and structure of the scanning array are described in Figure 3A. The series of oligonucleotides was made by coupling the bases through a mask to a polypropylene surface in the order in which they occur in the complement of the target sequence, starting from the first base at the 5′-ends. The 3′-ends of the tethered oligonucleotides are attached to the surface. The longest oligonucleotides were in the centre of the array and the tips of the large diamond shapes were occupied by mononucleotides. In between were cells containing all lengths between 1 and 23. The array as synthesised was symmetrical above and below the centre line of the template and each oligonucleotide was represented twice. The diamond-shaped template creates a series of chevron-shaped patches flanking the two ends of the array of diamond-shaped cells. Successive chevrons are occupied by oligonucleotides differing in length by 1 nt. The symmetrical duplication of the array above and below the centre line allowed us to compare and quantify the hybridisation yields of unmodified oligonucleotide **1** and $d^{4Et}C$-modified oligonucleotide **2** in a single hybridisation experiment. This was achieved by dividing the array along the center line with a thin rubber seal: one half was hybridised with oligonucleotide **1** and the other half with oligonucleotide **2** containing $d^{4Et}C$ (Fig. 3B). The hybridisation patterns for sequences **1** and **2** are similar but differ markedly in intensity. The signal from unmodified sequence **1** was barely detectable. Hybridisation signals of heteroduplexes formed between oligonucleotides **1** and **2** with the set of 23 immobilised probes of varying length were quantified and are shown in Figure 4.

The hybridisation results are in agreement with those obtained in solution. There are sharp transitions between regions that give high yield and those that give a yield close to the background. The 8, 9, 21, 22 and 23 nt oligonucleotides showed weaker hybridisation while the 10mers to 19mers produced much stronger hybridisation. The shortest oligonucleotide showing detectable hybridisation was an 8mer. Hybridisation with oligonucleotides shorter than 8mer was not detectable. The oligonucleotides showing maximum hybridisation were the 11mers to 15mers. For oligonucleotides longer than 15 nt the intensity of the hybridisation signals decreased progressively and merged with the background for the 23mer.

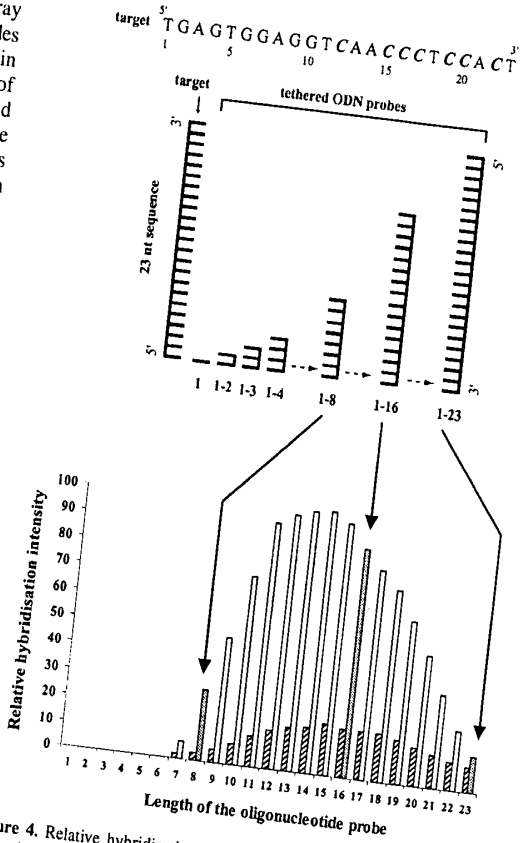


Figure 4. Relative hybridisation intensity of unmodified target sequence **1** (striped rectangle) and $d^{4Et}C$-modified target sequence **2** (open rectangle) as a function of the length of the tethered oligonucleotide probes varying in length from 1 to 23 nt. *C* = C or ^{4Et}C.

DISCUSSION

Hairpin structures are a common feature of single-stranded DNA and RNA molecules. These stable structures probably explain why some regions of targets are not accessible for heteroduplex formation with complementary oligonucleotides. Heteroduplex formation between a hairpin structure and a complementary oligonucleotide probe implies two main steps: opening of the structure to allow formation of one or a few base pairs in a transient nucleation complex followed by further base pairing between the two complementary strands to form a heteroduplex which is more stable than the starting structures. The secondary structure of nucleic acid targets can be minimised by using a hybridisation temperature above the melting temperature of the intramolecular structure or low salt concentration solutions. However, these conditions are not favourable for hybridisation of nucleic acid molecules with short oligodeoxynucleotides. We modified the target sequence to destabilise its secondary conformation to assist interactions with oligonucleotides probes.

We chose in this work to reduce the stability of the G:C base pair, which is the mainstay of hairpin structure. We chose to modify C because the preparation of analogues of C is simpler and easier than those of G. Among the modified nucleosides which fulfilled the criteria mentioned above, we chose $d^{4Et}C$, rather than *N*4-methyl-2′-deoxycytidine ($d^{4Me}C$), which has been used to suppress band compression in sequencing gels (21), because it has been shown previously that the destabilisation imparted to the G:C base pair by $d^{4Et}C$ appears to be much greater than the destabilisation due to $d^{4Me}C$ (22). As expected, replacement of dC with $d^{4Et}C$ in a 23mer oligodeoxynucleotide sequence which can form hairpin structure composed of a double-stranded stem of 8 nt and a 5 nt loop led to a decrease in the hairpin stability. The difference in stability between unmodified sequence **1** (T_m 70°C) and $d^{4Et}C$-modified sequence **2** made with $d^{4Et}C$ (T_m 40°C) is 30°C. Therefore,

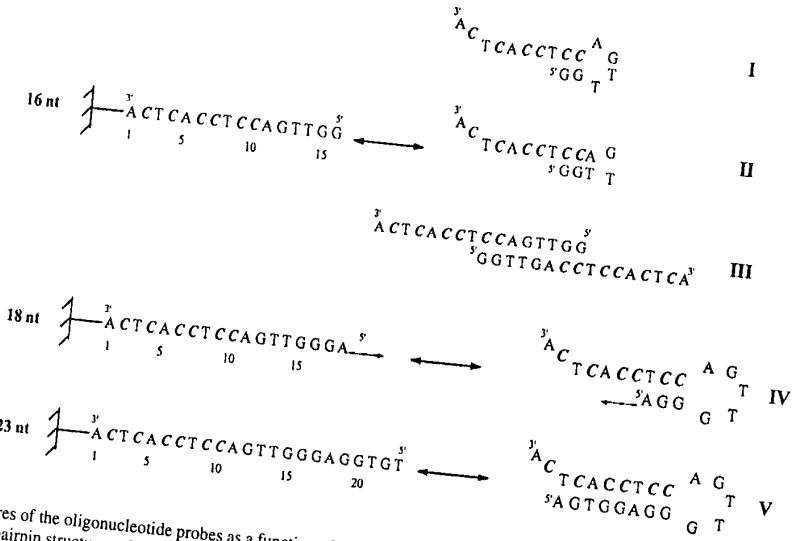


Figure 5. Possible structures of the oligonucleotide probes as a function of their length. Hairpin structures (I and II) and double-stranded structure (III) of the 16 nt oligomer probe. Possible hairpin structures of oligomer probes vary in length from 18 to 23 nt (IV and V). *C* = C or ^{4Et}C. All sequences are in the 'deoxy' series.

replacement of a G:C base pair in the sequence by a G:^{4Et}C base pair led to an average T_m decrease of ~6°C per substitution. This T_m decrease of 6°C is higher than the value of 4–5°C described in a previous study carried out with a set of nonamer intermolecular duplexes (16). The difference could be due to the particular hairpin structure of the sequence. The T_m decrease of the hairpin structure increased availability of the target for hybridisation with complementary probes. A gel shift experiment carried out with nonadeoxynucleotide probe **3** showed that ~10 equiv. were able to open most of the hairpin structure of modified sequence **2** involving d^{4Et}C and form heteroduplex. When the affinity of the probe was increased by adding the intercalating molecule 6-chloro-2-methoxyacridine, which improves binding due to the stacking interaction between the intercalator and the base pairs of the probe–target duplex, heteroduplex was formed with only equimolar quantities of probe.

In contrast to this result, only 30% of unmodified sequence **1** were involved in heteroduplex in the presence of 600 equiv. of probe **4**.

These results were verified by hybridisation to an oligonucleotide array. Under the same experimental conditions only d^{4Et}C-modified oligonucleotide **2** gave heteroduplexes with complementary oligonucleotide probes from 8mers to 23mers. The oligonucleotide probes giving maximum hybridisation are those ranging from 11 to 15 nt. Oligonucleotide probes longer than 16mer have the potential to form self-associated complexes, which may not be available for heteroduplex formation with the target. For example, 16mer oligonucleotide could adopt two hairpin structures: the first with a stem of two G:C base pairs and a loop of 4 nt (Fig. 5, structure I), the second containing a stem of 3 nt and a loop of only 2 nt (Fig. 5, structure II). However, these hairpin structures are not very stable and it seems more likely that the tethered 16 nt oligonucleotides are close enough to form intermolecular structure due to the high surface density. In this way they can form a double-stranded structure involving four G:C base pairs, two A:T pairs and two G:T mismatches (Fig. 5, structure III). The G:T mismatch is well known as among the more stable mismatch base pairs. Similarly, the 17 nt oligonucleotide could adopt an intermolecular structure involving one more G:T mismatch. Oligonucleotide probes from 18 to 23 nt can form hairpin structures involving a loop of 5 nt and a stem with a length of from 3 to 8 nt, respectively (Fig. 5, structures IV and V). This explains the decrease in hybridisation yield when the length of the probe increases.

There are other possibilities for target modification to produce or to augment the effects illustrated above. In the case of the G:C base pair we can increase the destabilisation by using bulkier groups (22). We can also decrease the stability by substitution of deoxyguanosine with 7-deaza-2′-deoxyguanosine (23). The stability of the A:T base pair could be reduced by substituting deoxythymidine with deoxyuridine and deoxyadenosine with 7-deaza-2′-deoxyadenosine (24,25), for example. Modifying the target in this way leads at the same time to a decrease in heteroduplex stability with oligonucleotide probes.

Other chemical modifications can be introduced in oligonucleotide probes to increase their affinity for the target. We have shown that tethering an intercalating agent, such as a 6-chloro-2-methoxyacridine molecule, at the end of the oligonucleotide probe enhanced target sequence hybridisation. Nucleoside analogues, such as *C*5-methyl- and *C*5-propynylpyrimidine nucleosides (26) or 2,6-diaminopurine 2′-nucleoside (27), known to have stabilising properties, can be also incorporated into oligonucleotide probes. In a preliminary experiment we showed that an oligonucleotide array including d^{5Me}C instead of dC, complementary to the tRNA^{Phe} sequence, substantially enhanced signal intensity with all sequences relative to the unmodified array.

CONCLUSIONS

Secondary structure in target sequences can prevent duplex formation with short oligonucleotides, leading to difficulties in the design and interpretation of experiments using oligonucleotide arrays. For the analysis of gene expression levels by hybridisation to gene-specific oligonucleotides, it is necessary to use a set of up to 20 different oligonucleotides to ensure that one or a few will give strong and specific interaction with each mRNA or cDNA target (2,8). For resequencing or genotyping there is less choice of probe; probes must embrace actual or potential variant sites and if there is no probe that gives good duplex yield because of folding in the target, the method cannot be used without modification. Base substitutions in the target and/or probe would be a simple, potentially general method to overcome the problem. The tests described in this paper show that a target with a very stable stem, which almost completely prevents hybridisation to complementary oligonucleotides, can be induced to hybridise by base substitutions which weaken secondary structure in the target. The modified base used in these studies (^{4Et}C) can be readily incorporated into targets by DNA polymerases, which should allow the method to be applied to all current applications of oligonucleotide arrays.

ACKNOWLEDGEMENTS

The authors wish to thank T. Fell for construction of the apparatus used in the array synthesis. We are grateful to J.K. Elder for data analysis, software and help with the computer and to F.-X. Barre, K. Mir, V. Regnier and M. Sohail for helpful discussions. H.-K.N. was supported by a TMR Marie Curie research grant.

REFERENCES

1. Duggan,D.J., Bittner,M., Chen,Y., Meltzer,P. and Trent,J.M. (1999) *Nature Genet.*, **21**, 10–14.
2. Lipshutz,R.J., Fodor,S.P.A., Gingeras,T.R. and Lockart,D.J. (1999) *Nature Genet.*, **21**, 20–44.
3. Bowtell,D.D.L. (1999) *Nature Genet.*, **21**, 25–32.
4. Hacia,J.G. (1999) *Nature Genet.*, **21**, 42–47.
5. Williams,J.C., Case-Green,S.C., Mir,K.U. and Southern,E.M. (1994) *Nucleic Acids Res.*, **22**, 1365–1367.
6. Milner,N., Mir,K.U. and Southern,E.M. (1997) *Nature Biotechnol.*, **15**, 537–541.
7. Southern,E.M., Mir,K. and Shchepinov,M. (1999) *Nature Genet.*, **21**, 5–9.
8. Lockhart *et al.* (1996) *Nature Biotechnol.*, **14**, 1675–1680.
9. Cronin,M.T., Fucini,V., Kim,S.M., Masino,R.S., Wespi,R.M. and Miyada,C.G. (1996) *Hum. Mutat.*, **7**, 244–255.
10. Weiler,J., Gausepohl,H., Hauser,N., Jensen,O.N. and Hoheisel,J.D. (1997) *Nucleic Acids Res.*, **25**, 2792–2799.
11. Thuong,N.T. and Asseline,U. (1991) In Eckstein,F. (ed.), *Oligonucleotides and Analogues: A Practical Approach*. IRL Press, Oxford, UK, pp. 283–308.
12. Matson,R.S., Rampal,J.B. and Coassin,P.J. (1994) *Anal. Biochem.*, **217**, 306–310.
13. Southern,E.M., Case-Green,S.C., Elder,J.K., Johnson,M., Mir,K.U., Wang,L. and Williams,J.C. (1994) *Nucleic Acids Res.*, **22**, 1368–1373.
14. Sohail,M. and Southern,E.M. (2000) *Mol. Cell Biol. Res. Commun.*, in press.
15. Elder,K.J., Johnson,M., Milner,N., Mir,K.U., Sohail,M. and Southern,E.M. (1999) In Schena,M. (ed.), *DNA Microarrays: A Practical Approach*. IRL Press, Oxford, UK, pp. 61–76.
16. Nguyen,H.-K., Bonfils,E., Auffray,P., Costaglioli,P., Schmitt,P., Asseline,U., Durand,M., Maurizot,J.-C., Dupret,D. and Thuong,N.T. (1998) *Nucleic Acids Res.*, **26**, 4249–4258.
17. Maniatis,T., Ptashne,M., Beckman,K., Kleid,D., Flashman,S., Jeffrey,A. and Maurer,R. (1975) *Cell*, **5**, 109–113.
18. Rosenberg,M. and Court,D. (1979) *Annu. Rev. Genet.*, **13**, 319–351.
19. Asseline,U., Delarue,M., Lancelot,G., Toulmé,J.-J., Thuong,N.T., Montenay-Garestier,T. and Hélène,C. (1984) *Proc. Natl Acad. Sci. USA*, **81**, 3297–3301.
20. Toulmé,J.-J., Krisch,H.M., Loreau,N., Thuong,N.T. and Hélène,C. (1986) *Proc. Natl Acad. Sci. USA*, **83**, 1227–1231.
21. Li,S., Haces,A., Stupar,L., Gebeyehu,G. and Pless,R. (1993) *Nucleic Acids Res.*, **21**, 2709–2714.
22. Nguyen,H.-K., Auffray,P., Asseline,U., Dupret,D. and Thuong,N.T. (1997) *Nucleic Acids Res.*, **25**, 3059–3065.
23. Seela,F., Tran-Thi,Q.-H. and Franzen,D. (1982) *Nucleic Acids Res.*, **21**, 4338–4343.
24. Ono,A., Ohdoi,C., Matsuda,A. and Udea,T. (1992) *Nucleosides Nucleotides*, **11**, 227–235.
25. Seela,F. and Grein,T. (1992) *Nucleic Acids Res.*, **20**, 2297–2306.
26. Sagi,J., Czuppon,A., Katjar,M., Szabolcs,A., Szembo,A. and Otwos,L. (1982) *Nucleic Acids Res.*, **10**, 6051–6066.
27. Chollet,A., Chollet-Damerius,A. and Khawashima,E.H. (1986) *Chem. Scr.*, **26**, 37–40.

Proc. Natl. Acad. Sci. USA
Vol. 94, pp. 2903–2908, April 1997
Biochemistry

Defining the chemical groups essential for *Tetrahymena* group I intron function by nucleotide analog interference mapping

SCOTT A. STROBEL† AND KANAKA SHETTY

Department of Biochemistry and Molecular Biophysics, 260 Whitney Avenue, Yale University, New Haven, CT 06520

Communicated by Thomas A. Steitz, Yale University, New Haven, CT, January 27, 1997 (received for review November 18, 1996)

ABSTRACT **Improved atomic resolution biochemical methods are needed to identify the chemical groups within an RNA that are essential to its activity. As a step toward this goal, we report the use of 5′-*O*-(1-thio)inosine monophosphate (IMPαS) in a nucleotide analog interference mapping (NAIM) assay that makes it possible to simultaneously, yet individually, determine the contribution of almost every N2 exocyclic amine of G within a large RNA. Using IMPαS, we identified the exocyclic amines that are essential for 5′ or 3′ exon ligation by the *Tetrahymena* group I intron. We report that the amino groups of three phylogenetically conserved guanosines (G111, G112, and G303) are important for 3′ exon ligation. The amine of G22, as well as the amines of the other four guanosines within the P1 helix, are essential for ligation of the 5′ exon. Previous work has shown that point mutation of either G22 or G303 to an adenosine (A) substantially reduces activity. Like inosine, adenosine lacks an N2 amino group. Interference rescue of the G22A and G303A point mutations was detected at the site of mutation by NAIM using 5′-*O*-(1-thio)diaminopurine riboside monophosphate (DMPαS), an adenosine analog that has an N2 exocyclic amine. The G22A point mutant could also be rescued by incorporation of DMPαS at A24. By analogy to genetics, there are interference phenotypes comparable to loss of function, reversion, and suppression. This method can be readily extended to other nucleotide analogs for the analysis of chemical groups essential to a variety of RNA and DNA activities.**

The *Tetrahymena* group I intron catalyzes two consecutive transesterification reactions in the course of RNA self-splicing (1). Splicing is initiated by nucleophilic attack at the 5′ splice site by the 3′-hydroxyl of an exogenous guanosine (G) cofactor. This results in covalent attachment of the G onto the intron and release of the 5′ exon. Previous to the second step of splicing, the RNA undergoes a conformational transition in which G414, the 3′ terminal nucleotide in the intron, enters the G binding site. This conformational rearrangement brings the exon boundaries into close proximity. In the second step of splicing, the 5′ exon attacks the 3′ splice site to produce ligated exons. The intron can also catalyze the reverse of these two reactions resulting in ligation of the exons onto the intron (2–4).

During the first step of splicing, the location of the 5′ splice site is defined by a phylogenetically conserved G·U wobble pair in the P1 helix (Fig. 1) (5–11). This noncanonical pair forms between a U at the end of 5′ exon and a G located in the internal guide sequence. The contributions made by specific chemical groups within the wobble pair were defined by systematically substituting the G·U with a series of nucleotide analogs, including inosine and diaminopurine riboside (10–12). Inosine is a G analog that lacks the N2 exocyclic amine, while diaminopurine riboside is an A analog with the amine. These experiments demonstrated that the exocyclic amine of G, when presented on the minor groove surface by the wobble configuration, contributes to tertiary interaction with the ribozyme active site, reaction fidelity, and stabilization of the chemical transition state (10, 11).

There are 106 other guanosines within the intron, including several that are phylogenetically conserved (13). It is reasonable to expect that the N2 amino groups of a subset of these are important for proper splicing activity, though there is currently no simple method available to identify them. With substantial effort, specific positions could be tested using a semisynthetic method to prepare full-length RNAs with a single inosine substitution (14). However, this approach is sufficiently labor intensive that only a few nucleotide positions could be tested. A more efficient method is required to identify the essential amines or other important chemical groups within an RNA.

Interference mapping by phosphorothioate substitution is an effective method to identify the nucleotide positions where replacing the nonbridging R_P oxygen with a sulfur is detrimental to activity (15–22). The sites of phosphorothioate incorporation are readily detected by cleavage of the phosphate linkage with iodine (15). This makes it possible to simultaneously, yet individually, quantitate every position within the RNA by an assay that is as simple as RNA sequencing. Krupp and coworkers (23, 24) extended the phosphorothioate interference approach to nucleotide analogs by chemically tagging 2′-deoxy and 2′-methoxy derivatives with a phosphorothioate marker. They used these derivatives to identify the 2′-hydroxyl groups in tRNA that are essential for recognition by RNase P.

In principle, the R_P phosphorothioate could be used as a general chemical tag for numerous nucleotide analogs. Development of a complete library of phosphorothioate-tagged analogs would provide a powerful atomic resolution tool for RNA structure/function studies. As a step toward this goal, we report the use inosine and diaminopurine riboside phosphorothioates to identify the N2 exocyclic amines of G that are essential for 5′ or 3′ exon ligation by the *Tetrahymena* group I intron. In combination with point mutagenesis, this pair of nucleotide derivatives makes it possible to detect essential amino groups by both inhibition and restoration of intron function.

METHODS

Synthesis of Phosphorothioate Analogs. 5′-*O*-(1-thio)-inosine triphosphate (ITPαS) was synthesized from inosine in a procedure analogous to that reported for the synthesis of

Abbreviations: NAIM, nucleotide analog interference mapping; IMPαS or ITPαS, 5′-*O*-(1-thio)inosine monophosphate or triphosphate; DMPαS or DTPαS, 5′-*O*-(1-thio)diaminopurine riboside monophosphate or triphosphate; GMPαS or AMPαS, 5′-*O*-(1-thio)guanosine or adenosine monophosphate; rT, 5-methyl uridine; dC, 2′-deoxy-cytidine; dT, thymidine; rT(-1)S, GGCCCUC(rT)AAAAA; dT(-1)S, CCCUC(dT)AAAAA; rT(-1)P, GGCCCUC(rT); dT(-1)P, CCCUC(dT); dC(-1)P, CCCUC(dC).

†To whom reprint requests should be addressed. e-mail: strobel@csb.yale.edu.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

0027-8424/97/942903-6$2.00/0
PNAS is available online at **http://www.pnas.org**.

5′-*O*-(1-thio)-2′-deoxyadenosine triphosphate (25). Inosine (100 mg, 0.37 mmol) was dried under vacuum at 110 °C for 12 h and dissolved in triethylphosphate (5 ml). The nucleoside was reacted with thiophosphorylchloride (42 μl, 0.41 mmol, 1.1 equivalents) in the presence of trioctylamine (180 μl, 0.41 mmol, 1.1 equivalents) at 0°C for 30 min. It was converted directly to the triphosphate by addition of tributylammonium pyrophosphate (0.5 g, 1.0 mmol) in triethylphosphate (3 ml). The product was precipitated by the addition of excess triethylamine (2.5 ml), centrifuged, and decanted. The residue was dissolved in aqueous triethylammonium bicarbonate (10 ml, 1.0 M, pH 7.5). Purification by DEAE-A25 Sephadex chromatography using a triethylammonium bicarbonate buffer gradient (0.05–0.8 M) afforded ITPαS as a diastereomeric mixture in 23% yield. ^{31}P NMR (H_2O): 43.96 (*m*), −10.21 (*d*), −24.13 (*t*). UV: 250 nm λ_{max}. A similar procedure afforded 5′-*O*-(1-thio)diaminopurine ribosyl triphosphate (DTPαS) in 26% yield. ^{31}P NMR (H_2O): 42.61 (*m*), −6.24 (*d*), −23.18 (*t*). UV: bimodal with a λ_{max} at 256 and 280 nm.

Plasmids. Plasmids pUCL-21G414R and pUCL+1 were generated from plasmid pT7L-21 by PCR amplification with oligonucleotides containing the appropriate sequence changes. The pUCL-21G414R plasmid contains a new *Ear*I site for use in run-off transcription to produce an RNA with G414 as the terminal nucleotide. pUCL+1 replaces the first 21 nt of the wild-type intron with the sequence 5′-GAAUUACACAAUUAUACCAUU. The sequence contains mutations (underlined) that eliminate the circularization sites present within the wild-type L1 loop (26). This sequence includes a G as the first base of the intron which is analogous to the exogenous G added after the first step of splicing (27). pUCL-21G414R G303A and pUCL+1 G22A are single point mutations of the parent plasmids.

Transcription Reactions. L-21 G^{414} RNA (Fig. 1) was transcribed from *Ear*I digested pUCL-21G414R by T7 RNA polymerase using 1 mM of each NTP. L+1 *Sca*I RNA was transcribed from *Sca*I-digested pUCL+1 under the same conditions. 5′-*O*-(1-thio)inosine monophosphate IMPαS or DMPαS were randomly incorporated into the RNA transcripts at a level of ≈5 per molecule (5%) using 0.4 mM ITPαS or 25 μM DTPαS in the transcription mixture, respectively. RNAs were also prepared with ≈5% GMPαS or AMPαS using 50 μM

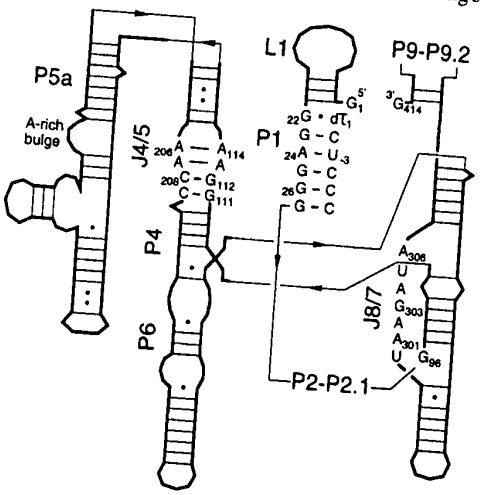


FIG. 1. Composite diagram of the *Tetrahymena* group I intron. The L-21 G^{414} RNA begins at G22 and ends at G414, but does not include the L1 loop. The L+1 *Sca*I RNA begins at G1, includes a mutated version of the L1 loop, and terminates 5 nt before G414. The composite RNA is shown bound to the oligonucleotide dT(-1)P which is the substrate for the reaction with L+1 *Sca*I. The diagram also identifies the nucleotide positions, and the duplex and single-stranded regions discussed in the text.

of the S_P form of the triphosphate (18). Levels of IMPαS and DMPαS incorporation were determined by normalizing to transcripts containing 5% GMPαS and AMPαS, respectively. There was no detectable difference in the interference pattern at any of the levels of phosphorothioate incorporation tested (from 0.5% to 10%). A level of 5% incorporation provided an easily detected signal, but remained low enough to avoid interference between multiple substitutions within a single RNA. Following transcription, the RNAs were purified by PAGE and eluted into 10 mM Tris·HCl, pH 7.5/1 mM EDTA (TE buffer). The RNAs were ethanol precipitated and resuspended in TE buffer.

A Ribozyme with Inosine in Place of Guanosine. An L-21 *Sca*I form of the *Tetrahymena* ribozyme was prepared with inosine in place of G at all positions except the 5′ terminal nucleotide, G22. *Sca*I-digested pT7L-21 plasmid was used as a template for T7 RNA polymerase transcription with 4 mM inosine triphosphate and 10 mM GMP in the transcription mixture.

Radiolabeling. Oligonucleotides rT(-1)S [GGCCCUC(rT)AAAAA] and dT(-1)S [CCCUC(dT)AAAAA] were 3′ end labeled with [α-^{32}P]cordycepin using poly(A) polymerase. Oligonucleotides rT(-1)P [GGCCCUC(rT)], dT(-1)P [CCCUC(dT)], and dC(-1)P [CCCUC(dC)] were 5′ end labeled with [γ-^{32}P]ATP using polynucleotide kinase. All oligonucleotides were purified on a native polyacrylamide gel prior to use and eluted into TE buffer.

The L-21 G^{414} RNA was treated with alkaline phosphatase and 5′ end-labeled with [γ-^{32}P]ATP using polynucleotide kinase. The radiolabeled 5′ control RNAs were gel purified and eluted into TE buffer.

Reaction Conditions. The L+1 *Sca*I and L-21G^{414} RNAs perform RNA ligation reactions that are analogous to the reverse of the first and second steps of RNA splicing, respectively. Permissive and nonpermissive reaction conditions were chosen for nucleotide analog interference mapping (NAIM). Permissive conditions are defined as 50 nM enzyme in 20 mM $MgCl_2$, 2 mM $Mn(OAc)_2$, and 50 mM Hepes (pH 7.0) for 10 min at 50°C with 5 nM of an all ribose oligonucleotide [rT(-1)S for L-21G^{414} and rT(-1)P for L+1 *Sca*I RNA]. Nonpermissive conditions are defined as 50 nM enzyme in 4 mM $MgCl_2$, and 50 mM Hepes (pH 7.0) for 10 min at 50°C with an oligonucleotide (5 nM) containing a single 2′-deoxy substitution at the reaction site [dT(-1)S for L-21G^{414} and dT(-1)P for L+1 *Sca*I RNA]. The single deoxy substitution does not affect oligonucleotide binding, but it does reduce the rate of the chemical reaction by at least a 1000-fold (28). RNAs were incubated at 50°C in the appropriate reaction buffer for 10 min prior to the addition of the oligonucleotide. The reaction was quenched by the addition of two volumes of stop solution (8 M urea/50 mM EDTA/0.01% bromophenol blue/0.01% xylene cyanol). The phosphorothioates were cleaved by the addition of 1/10th volume of 100 mM iodine in ethanol (15). The solution was heated to 90°C for 2 min and the cleavage products resolved by electrophoresis on a 5% polyacrylamide gel. The intensities of individual bands were quantitated by PhosphorImager (Molecular Dynamics) analysis.

RESULTS

A Group I Intron Containing Inosine in Place of Guanosine Is Inactive. Previous work has shown that the N2 exocyclic amine of G22 is important for the first step of group I intron activity (10, 11). We set out to determine if any other N2 amino groups of G play an important role in the activity. To address this question, a variation of the L-21 *Sca*I form of the intron was prepared in which every G except G22 was replaced with inosine. The inosine ribozyme demonstrated no activity under any reaction conditions when incubated for several hours with the oligonucleotide substrate rT(-1)S. This included incubation of the substrate with 1 μM ribozyme, 10 mM GMP, and 100 mM Mg^{2+} ion. Failure to detect activity under these conditions demonstrates that N2 amino groups at positions

other than G22 are important. The locations of the essential groups could not be determined from this assay.

NAIM. We employed NAIM to identify the location of the essential N2 amino groups. The method involves four steps: (*i*) the phosphorothioate derivative of the nucleotide analog is randomly incorporated into the RNA by *in vitro* transcription, (*ii*) the active variants in the RNA population are selected from the inactive transcripts, (*iii*) the phosphorothioate linkage is cleaved by treatment with iodine, and (*iv*) the sites of analog substitution that are detrimental to RNA activity are identified by gel electrophoresis and autoradiography (Fig. 2). Gaps in the sequencing ladder define the sites where analog substitution is detrimental to activity. By this approach it is possible to simultaneously, yet individually, test the importance of an inosine substitution at every G within the group I intron.

Defining the N2 Exocyclic Amines Important for 3′ Exon Ligation. Variants of the L-21 G^{414} RNA containing ≈5 IMPαS substitutions per molecule were reacted with the 3′ end-labeled oligonucleotide substrate dT(-1)S under nonpermissive conditions (Fig. 2) (4). In this reaction, the 3′ end of the oligonucleotide (AAAAAA*), which is analogous to the 3′ exon, is transferred from the substrate onto the 3′ end of the active variants in the intron population (3). Inactive RNAs are not radiolabeled.

The intensity of the iodine sequencing ladder at three nucleotide positions is substantially reduced among the active RNA variants containing IMPαS compared with the GMPαS control (Fig. 3). This includes G111, G112, and G303. There is a 3-fold phosphorothioate effect at G303, but IMPαS is reduced an additional 8-fold relative to GMPαS at this position (Fig. 3*A* lanes 1 and 2). The 3- and 5-fold effects were observed upon substituting inosine at G111 and G112, respectively. There was no phosphorothioate effect at these positions (Fig. 3*B*). Normal cleavage intensities at G111, G112, and G303 were observed among the total population of RNAs containing IMPαS, as assayed by direct 5′ end labeling (Fig. 3*A*, lanes 7 and 8; Fig. 3*B*). This confirms that the loss of signal does not result from a failure of the polymerase to incorporate IMPαS into the RNA at these sites. This is further supported by the observation that the cleavage intensities at G111, G112, and G303 are partially recovered when the RNAs were reacted with the oligonucleotide rT(-1)S under permissive conditions (Fig. 3*A*, lane 5 and 6). There was no position where IMPαS substitution enhanced activity.

A total of 93 of the 106 guanines in the group I intron could be characterized in this assay. A strong phosphorothioate effect at G96 resulted in a band intensity that was too low in the GMPαS RNA for the IMPαS RNA to be informative. Other positions could not be accurately measured because the cleavage products were not be sufficiently resolved from the full-length ligated intron.

Defining the N2 Exocyclic Amines Important for 5′ Exon Ligation. To determine if 5′ exon ligation has the same N2 amino group requirements as ligation of the 3′ exon, variants of the L+1 *Sca*I RNA containing IMPαS substitutions were reacted with the radioactively labeled oligonucleotide dT(-1)P. In this reaction, the 5′ exon is covalently transferred onto the 5′ end of the active introns with displacement of the terminal guanosine (G1) (2).

For the 5′ splicing assay, the iodine sequencing ladder was substantially reduced at all five Gs in the internal guide sequence (Fig. 3*B*). No phosphorothioate effect was measured at any of these positions, so the reported effects are due entirely to inosine substitution. The strongest effect was at G22 where the magnitude of the interference was greater than 10-fold. This is the expected result based upon previous studies of the G·U wobble pair, and reconfirms that the exocyclic amine of G22 is essential for the activity of the intron (10, 11). Smaller effects (between 2- and 5-fold) were observed at other Gs within the P1 helix, including G23, G25, G26, and G27. If the L+1 *Sca*I IMPαS RNA was reacted with rT(-1)P under permissive conditions, cleavage bands were detected for all of the internal guide sequence

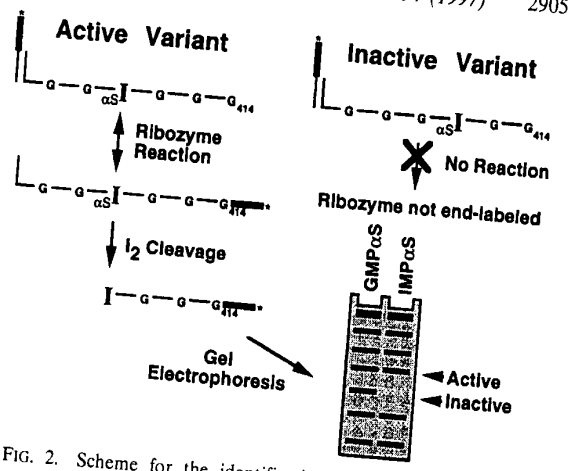


FIG. 2. Scheme for the identification of the chemical groups important for RNA activity by NAIM. In this example, IMPαS is randomly incorporated in place of G at a low ratio into the L-21 G^{414} RNA. Active variants covalently transfer the 3′ end of the exon onto the 3′ terminus of the intron. Hydrolysis of the phosphorothioate linkage with iodine breaks the RNA at the sites of IMPαS incorporation. Separation of the cleavage fragments by PAGE produces a sequencing ladder that represents the sites tolerant of IMPαS substitution. In contrast, inactive variants fail to react with the oligonucleotide which leaves a gap in the sequencing ladder. RNAs containing GMPαS are tested to insure that loss of activity is not due to the phosphorothioate group used to chemically tag the inosine.

residues, confirming that the loss of signal is due to loss of activity, rather than a failure of the RNA polymerase to incorporate IMPαS into the RNA at these sites.

The three positions identified as important for the 3′ exon ligation reaction were tolerated in the 5′ ligation reaction. This included G303, which in the 3′ ligation reaction displayed neither a phosphorothioate effect nor an effect from inosine substitution. The reaction also tolerated a phosphorothioate substitution at G96. This does not appear to be due to a general lack of selectivity in the L+1 *Sca*I reaction. Several of the phosphorothioate effects at adjacent positions, including A302, A304, U305, and A306, remained or were enhanced for this reaction (data not shown).

Interference Rescue of the G303A Mutation. The strongest IMPαS interference signal for the L-21 G^{414} reaction was at G303. Mutation of G303 to an A reduces the activity of the L-21 *Sca*I ribozyme 30-fold (29). Based upon the interference results above, this can be at least partially explained by the loss of the exocyclic amine upon mutation of G303 to an A. Diaminopurine riboside (2-amino adenosine; abbreviated as D) is an analog of adenine that contains an exocyclic amine at the N2 position of the purine ring comparable to G. We postulated that if the amine at G303 is essential for catalytic function, then activity could be rescued by repeating the interference experiment in the G303A mutant background using the nucleotide analog DMPαS (Fig. 4*A*).

There are two ways that DMPαS could rescue activity. The first, and most likely outcome, is that DMPαS could provide an exocyclic amine at the position originally mutated to an adenosine. This would appear on the interference sequencing gel as a single band of strong intensity at the same nucleotide position that is missing in the IMPαS interference experiment. Second, DMPαS could rescue activity at a position other than the original site of mutation. Both of these possibilities could be detected in a single interference mapping experiment.

Variants of the L-21 G^{414} G303A RNA containing approximately five random AMPαS or DMPαS substitutions per molecule were reacted with the 3′ end-labeled oligonucleotide rT(-1)S. The resulting interference pattern is shown in Fig. 4*B*. A very weak signal was detected at position A303 in the RNA

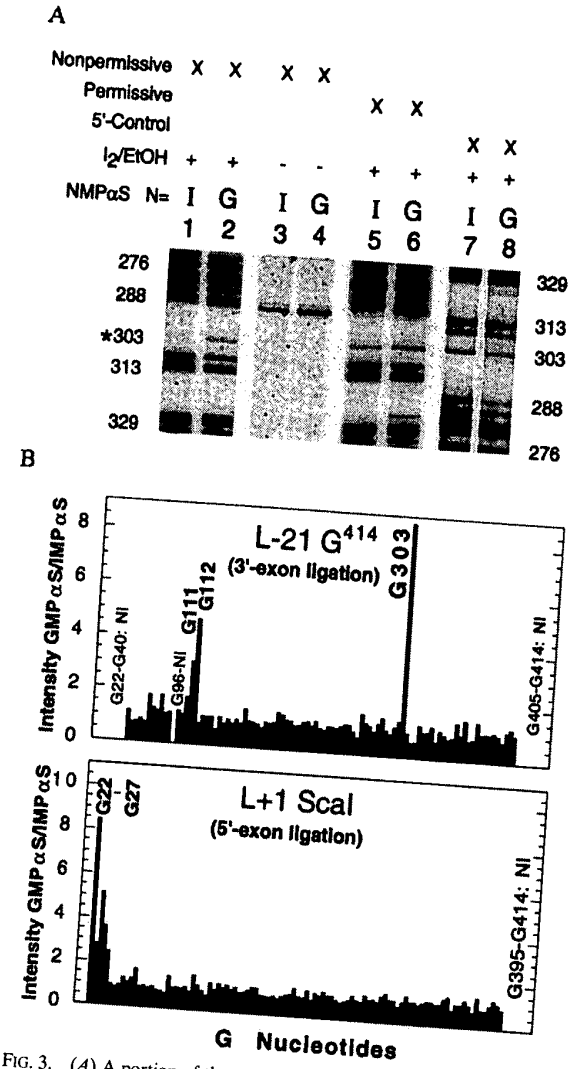


FIG. 3. (*A*) A portion of the sequencing autoradiogram showing the reaction of L-21 G^{414} IMPαS or GMPαS RNA with dT(-1)S under nonpermissive conditions (lanes 1–4) or rT(-1)S under permissive conditions (lanes 5 and 6). The nucleotide positions correlating to the cleavage products are shown to the left for lanes 1–6 and to the right for lanes 7 and 8. For the 5′ control (lanes 7 and 8), the L-21 G^{414} is enzymatically labeled at the 5′ end. This results in a reversal of the sequence orientation on the gel, but shows the actual level of analog incorporation at all nucleotide positions. RNA degradation independent of iodine treatment is consistently observed at A290. This is clearly seen in lanes 3 and 4. (*B*) Histogram showing the ratio of GMPαS intensity divided by IMPαS intensity for each position in both the L-21 G^{414} and L+1 *ScaI* RNAs when reacted under nonpermissive conditions and normalized to the actual levels of incorporation as defined by the 5′ control. Intensities were measured by PhosphorImager analysis. A value of 1 indicates that an inosine substitution at that site has no effect. A value greater than 2 indicates that inosine is detrimental, while a value less than 0.5 (not detected) would indicate that inosine substitution is advantageous at that site. Positions G22–G40 could not be accurately measured in the 5′ control, so these values were normalized using the intensity of each band under permissive reaction conditions. Thus, the magnitude of the effect at G22–G27 is a minimum estimate. All values were measured at least three times in separate experiments and are accurate to within 25% error. Sites that were not informative in the assay are indicated on the graph with the initials NI.

containing AMPαS (lane 2). The A303 signal is restored more than 5-fold in the RNA containing DMPαS (lane 1). This indicates that the activity of the G303A mutant can be at least partially restored by substituting an N2 amino group in the context of an adenosine nucleotide derivative, and confirms the importance of this chemical group to intron activity.

Interference Rescue of a G22A Mutation. The strongest site of IMPαS interference for the L+1 *ScaI* RNA was at G22. G22 forms a wobble pair with the U at position −1 in the oligonucleotide. Mutation of G22 to an A reduces the second order rate constant for cleavage by the L-21 *ScaI* form of the ribozyme by almost 500-fold (11). This is due to the loss of the exocyclic amine at G22, and the conversion of the G·U wobble pair to an A-U Watson–Crick pair. We set out to determine if the G22A mutation could also be rescued by interference with DMPαS.

Diaminopurine forms a stable Watson–Crick pair with U and a relatively unstable wobble pair with C (30). We measured the extent of DMPαS rescue using oligonucleotide substrates with either a U or a C at position −1 (Fig. 5). Activity was rescued more than 10-fold (compared with an RNA containing AMPαS and normalized for the extent of reaction and the level of analog incorporation) at position A22 when L+1 *ScaI* G22A RNA containing DMPαS was reacted with dC(-1)P. Only a 3-fold rescue was observed when the RNA was reacted with dT(-1)P. The DMPαS interference rescue experiment confirms the IMPαS result, thus showing by both positive and negative interference that the amine of G22 is important for 5′ exon ligation. The requirement for a mutation at a second site (D·C gave better rescue than D-U) confirms that the exocyclic amine makes its greatest contribution to splicing in the context of a wobble pair, where the amino group is free to make a tertiary contact with the intron active site (10, 11).

The intensity of iodine cleavage at positions other than A22 was much lower in the G22A experiment than for the G303A experiment because there was not phosphorothioate effect at position 22. This made it possible to detect interference rescue at positions other than the site of mutation. In addition to the strong cleavage band at position 22, a second band of moderately strong intensity was present at A24, the only other A nucleotide within the P1 helix (Fig. 5*B*). This band is 5-fold more intense (after correcting for the extent of reaction and analog incorporation) in the DMPαS lane than in the AMPαS lane when the RNAs were reacted with dC(-1)P (Fig. 5*B*, lanes 1 and 2). There was a 3-fold rescue at A24 when the RNA is reacted with dT(-1)P (lanes 5 and 6). Close inspection of cleavage intensities at other positions within the RNA failed to detect other examples of interference rescue.

DISCUSSION

Nucleotide analog interference mapping is a general method to define the chemical basis of RNA function. In this example, we have shown that inosine can be incorporated into the RNA transcript as a 5′-phosphorothioate in place of guanosine. This simultaneously replaces the N2 amino group of G with a proton and the 5′ nonbridging R_P phosphate oxygen with a sulfur. The ability to cleave with iodine exclusively at the phosphorothioate linkage makes it possible to map the sites of inosine incorporation that are detrimental to RNA function.

In many ways the 5′-α-phosphorothioate is an ideal chemical tag. Sulfur has only a slightly larger van der Waals radius than oxygen, and the length of the P—S bond is only slightly longer than the P—O bond (31). For these reasons, the sulfur substitution is not expected to substantially distort the RNA secondary structure. The most noticeable effect upon RNA structure is a change in the ability of the phosphate group to coordinate metal ions (32); however, this can be compensated by the inclusion of thiophilic metals such as Mn^{2+} or Cd^{2+} in the reaction mixture (33, 34). Thus, for a large majority of nucleotides within an RNA, the phosphorothioate substitution is functionally silent and can serve as an effective tag for analog incorporation. For example, in this experiment, only 1 of 94 positions was not informative because of phosphorothioate interference, and that site was only uninformative for the 3′ exon ligation reaction.

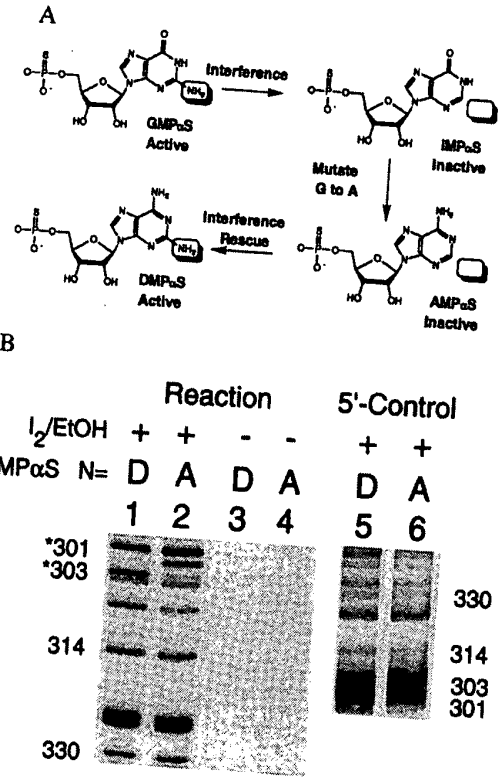


FIG. 4. (*A*) Scheme for interference rescue. Ribozyme activity is reduced by mutating a G that is intolerant of IMPαS substitution to an A. Activity is rescued by substituting the A with DMPαS. (*B*) A portion of the sequencing autoradiogram showing the reaction of L-21 G^{414} G303A DMPαS or AMPαS with rT(-1)S for 10 min at 50°C in 4 mM $MgCl_2$. The nucleotide positions correlating to the cleavage products are shown to the left for lanes 1–4 and to the right for lanes 5 and 6. For the 5′ control (lanes 5 and 6), the L-21 G^{414} RNAs are enzymatically labeled at the 5′ end. This results in a reversal of the sequence orientation on the gel, but shows the actual level of analog incorporation at all nucleotide positions. A visual comparison between the reaction and the 5′ control is complicated by strong phosphorothioate interference at A302 and A306. These bands are absent in both the DMPαS and AMPαS reaction lanes, but they are present in both 5′ controls.

Genetic Analogies. It might be useful to consider this approach using the language of genetics. Each nucleotide position within the RNA has a phenotype that is expressed as a band intensity on a sequencing gel. Reduction of signal intensity among the active fraction of RNA molecules under nonpermissive conditions (low metal and a 2′-deoxy substitution at the cleavage site) is analogous to a loss of function phenotype, and defines the positions where analog substitution is detrimental to RNA function. Corollaries of genetic revertants and suppressors can be generated by making a deleterious mutation in the RNA and screening for interference rescue. Restoration of activity by analog substitution at the site of mutation is analogous to a genetic revertant, while rescue by substitution at a distant position is a comparable to a genetic suppressor.

We observed at least one example of each of these phenotypes during our efforts to map the *Tetrahymena* group I intron. The N2 exocyclic amines of G111, G112, and G303 are important for 3′ exon ligation, but are not essential for ligation of the 5′ exon. All three of these nucleotides are located within the catalytic core of the intron. G22, G23, G25, G26, and G27 are important for 5′ exon ligation, but could not be measured for the 3′ exon ligation reaction. These five nucleotides are all in the P1 helix of the intron. Thus, there are eight examples of a loss of function phenotype. Mutation of either G22 or G303 to an A is deleterious, but activity could be restored by incorporating DMPαS at the sites of mutation. This provides two examples of a revertant. The activity of the G22A mutation could also be rescued by incorporating DMPαS at A24. This is an example of a suppressor. Given the parallels between nucleotide analog interference mapping and genetics, we feel that the term "chemogenetics," originally coined by Sigler, provides a convenient and descriptive one-word summary of the approach (35).

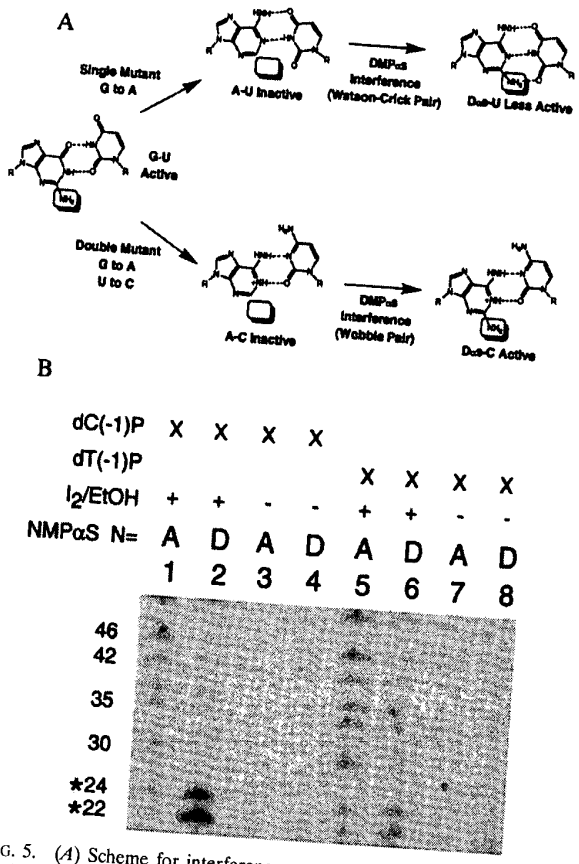


FIG. 5. (*A*) Scheme for interference rescue that requires a second mutation for maximum activity. Unlike the scheme in Fig. 5, activity of the G to A point mutant is not significantly rescued by DMPαS substitution unless a second nucleotide is also changed. In this case, U −1 must be mutated to a C. This restores the wobble configuration of the base pair. (*B*) A portion of the sequencing autoradiogram showing the reaction of L+1 *ScaI* DMPαS or AMPαS at 50°C in 4 mM $MgCl_2$. Reactions with dT(-1)P and dC(-1)P were for 2 and 10 min, respectively. The magnitude of the interference rescue was estimated by comparing the intensity of each band in the DMPαS lane to the intensity in the AMPαS lane and normalizing to the intensity at all other positions in the RNA.

Possible Roles for the Essential N2 Amino Groups. All of the amino groups identified as important in the exon ligation assays either stabilize the P1 duplex or are in close proximity to nucleotides known to interact with it. The site showing the highest level of interference in the second step of splicing is G303. It is located in the J8/7 region of the active site and is 94% conserved among all group I introns (439 total examples) (13). It is 100% conserved among the IC subgroup of introns which includes *Tetrahymena* (243 examples). This high level of conservation suggests that G303 plays an important role in intron function. The neighboring base, A302, has been shown to make a long-range hydrogen bonding contact with an important 2′-hydroxyl in the P1 helix (29). This tertiary contact is thought to position the 5′ exon into the intron active site for splicing. It is reasonable to expect that the exocyclic amine of G303 is also proximal to the P1 helix. It could contribute to

activity by making a direct hydrogen bond to P1, as proposed in the original model of Michel and Westhof (36), or the contribution could be indirect. For example, the amine might be essential for orienting A302 so that it can contact P1.

There are several possible explanations for the contribution of the G111 and G112 amines. G111 and G112 are both in the P4 region of the intron active site, and both form canonical Watson–Crick pairs with C209 and C208, respectively (37). The amino groups are located in the minor groove of the A-form RNA duplex (37). G112 is 89% conserved among all group I introns and 98% conserved among introns in subgroup IC. G111 is only 58% conserved overall, but it is 92% conserved among the subgroup IC1 and IC2 introns (13). This suggests that the exocyclic amines of G111 and G112 are essential for the *Tetrahymena* intron, but their contribution might not be universal.

The P4 helix is located within a densely packed region of the active site. The major groove face of G111 interacts with U305 in the J8/7 region (38). Photo-crosslinking demonstrates that the G·U of the P1 helix is positioned adjacent to the sheared A-A pairs that are coaxially stacked immediately above the G112-C208 pair (37, 39). The A-rich bulge in P5a docks into the minor groove of the P4 helix (37). Thus, G111 and G112 are in a central core of the intron, close to the J8/7 single-stranded region, the P1 helix, and the A-rich bulge in P5a.

The most trivial explanation for the detrimental effect of inosine substitution at G111 and G112 is that the exocyclic amines stabilize base pairing within the P4 helix. Although an I-C pair is less stable than a G-C pair (30), it seems unlikely that this is the explanation, given the high level of phylogenetic conservation at these two positions. It is more likely that these two amines are important for long-range tertiary interactions between other regions of the active site. The P1 helix is the most likely candidate due to the close proximity of the P1 helix to the minor groove of G111 and G112.

All of the N2 amino groups in the P1 helix are important for 5′ exon ligation including those at positions G22, G23, G25, G26, and G27. Except for G22, the sequence at these positions is not phylogenetically conserved, so the exocyclic amines are unlikely to be involved in essential tertiary interactions. The effects at G23–G27 most likely result from a reduction in P1 duplex stability. Improved P1 stability is also the most probable explanation for suppression of the G22A mutation by DMPαS at A24. This creates a D-U pair which is more stable than the wild-type A-U pair (30).

Unlike the other positions in the P1 helix, G22 is 100% conserved among all group I introns and forms a wobble pair with −1 U which is 98.7% conserved (13). The effect at G22 is consistent with previous experiments that characterized an RNA containing a single inosine substitution at position 22 (10, 11).

The crystal structure of the P4–P6 domain suggests that other N2 amino groups could be important for activity (37). These were not detected. It is possible that the P4–P6 domain and other regions within the intron are sufficiently stable and structurally redundant that folding is not impaired by the loss of a single hydrogen bonding group (40).

Conformational Changes in the Active Site Between the First and Second Steps of Splicing? It is uncertain why the G111, G112, and G303 amino groups are only essential for 3′ exon ligation. One intriguing possibility is that the difference in the interference pattern for G111, G112, and G303 reflects structural changes in the active site between the two steps of splicing. While this cannot be proven from the data presented here, there is a precedent for this interpretation. The phosphorothioate interference patterns for 5′ exon cleavage and 3′ exon hydrolysis are similar, but not equal (16, 18, 19). Many of the differences are in regions close to G303 and G112. Prior reports have suggested that the identity of the G22·U-1 pair at the 5′ exon cleavage site is only essential for the first step of splicing (7); while covariation between U305 and G111 is essential for 3′ splice site hydrolysis, but not for 5′ exon cleavage (38). Furthermore, methylation interference experiments on the group I intron sunY showed a different interference pattern for 5′ compared with 3′ exon hydrolysis (41). The difference in the inosine interference pattern, provides additional evidence that the conformation of the intron active site (specifically in areas proximal to the P1 helix) is not equivalent for the first and second steps of splicing.

Other Nucleotide Analogs and Other Nucleic Acid Polymers. In principle, this approach can be extended to several other nucleotide analogs and to any RNA or DNA polymer that can be enzymatically synthesized and functionally selected. "Functions" might be defined as catalysis, protein binding, secondary or tertiary structure formation, ligand binding, conformational heterogeneity, etc. NAIM will make it possible to identify chemical groups essential for a wide variety of RNA and DNA activities.

We thank Leo Beigelman and Jasenka Adamic for advice on phosphorothioate synthesis, Robin Gutell for performing phylogenetic comparisons, and Lara Weinstein, Jamie Cate, Sean Ryder, and Soumitra Basu for critical reading of the manuscript. This work was supported by National Institutes of Health Grant GM54839 and a Junior Faculty Research Award from the American Cancer Society to S.A.S.

1. Cech, T. R. (1990) *Annu. Rev. Biochem.* **59,** 543–568.
2. Green, R., Ellington, A. D. & Szostak, J. W. (1990) *Nature (London)* **347,** 406–408.
3. Beaudry, A. A. & Joyce, G. F. (1992) *Science* **257,** 635–641.
4. Mei, R. & Herschlag, D. (1996) *Biochemistry* **35,** 5796–5809.
5. Waring, R. B., Towner, P., Minter, S. J. & Davies, R. W. (1986) *Nature (London)* **321,** 133–139.
6. Doudna, J. A., Cormack, B. P. & Szostak, J. W. (1989) *Proc. Natl. Acad. Sci. USA* **86,** 7402–7406.
7. Barfod, E. T. & Cech, T. R. (1989) *Mol. Cell. Biol.* **9,** 3657–3666.
8. Pyle, A. M., Moran, S., Strobel, S. A., Chapman, T., Turner, D. H. & Cech, T. R. (1994) *Biochemistry* **33,** 13856–13863.
9. Knitt, D. S., Narlikar, G. J. & Herschlag, D. (1994) *Biochemistry* **33,** 13864–13879.
10. Strobel, S. A. & Cech, T. R. (1995) *Science* **267,** 675–679.
11. Strobel, S. A. & Cech, T. R. (1996) *Biochemistry* **35,** 1201–1211.
12. Green, R., Szostak, J. W., Benner, S. A., Rich, A. & Usman, N. (1991) *Nucleic Acids Res.* **19,** 4161–4166.
13. Damberger, S. H. & Gutell, R. R. (1994) *Nucleic Acids Res.* **22,** 3508–3510.
14. Moore, M. J. & Sharp, P. A. (1992) *Science* **256,** 992–997.
15. Gish, G. & Eckstein, F. (1988) *Science* **240,** 1520–1522.
16. Waring, R. B. (1989) *Nucleic Acids Res.* **17,** 10281–10293.
17. Ruffner, D. E. & Uhlenbeck, O. C. (1990) *Nucleic Acids Res.* **18,** 6025–6029.
18. Christian, E. L. & Yarus, M. (1992) *J. Mol. Biol.* **228,** 743–758.
19. Christian, E. L. & Yarus, M. (1993) *Biochemistry* **32,** 4475–4480.
20. Jeong, Y.-H., Kumar, P. K. R., Suh, Y.-A., Taira, K. & Nishikawa, S. (1994) *Nucleic Acids Res.* **22,** 3722–3727.
21. Hardt, W.-D., Warnecke, J. M., Erdmann, V. A. & Hartmann, R. K. (1995) *EMBO J.* **14,** 2935–2944.
22. Yu, Y.-T., Maroney, P. A., Darzynkiewicz, E. & Nilsen, T. W. (1995) *RNA* **1,** 46–54.
23. Gaur, R. K. & Krupp, G. (1993) *Nucleic Acids Res.* **21,** 21–26.
24. Conrad, F., Hanne, A., Gaur, R. K. & Krupp, G. (1995) *Nucleic Acids Res.* **23,** 1845–1853.
25. Arabshahi, A. & Frey, P. A. (1994) *Biochem. Biophys. Res. Commun.* **204,** 150–155.
26. Been, M. D. & Cech, T. R. (1987) *Cell* **50,** 951–961.
27. Cech, T. R., Zaug, A. J. & Grabowski, P. J. (1981) *Cell* **27,** 487–496.
28. Herschlag, D., Eckstein, F. & Cech, T. R. (1993) *Biochemistry* **32,** 8312–8321.
29. Pyle, A. M., Murphy, F. L. & Cech, T. R. (1992) *Nature (London)* **358,** 123–128.
30. Strobel, S. A., Cech, T. R., Usman, N. & Beigelman, L. (1994) *Biochemistry* **33,** 13824–13853.
31. Frey, P. A. & Sammons, P. (1985) *Science* **228,** 541–545.
32. Pecoraro, V. L., Hermes, J. D. & Cleland, W. W. (1984) *Biochemistry* **23,** 5262–5271.
33. Dahm, S. C. & Uhlenbeck, O. C. (1991) *Biochemistry* **30,** 9464–9469.
34. Piccirilli, J. A., Vyle, J. S., Caruthers, M. H. & Cech, T. R. (1993) *Nature (London)* **362,** 85–88.
35. Sigler, P. B. (1994) *Nat. Struct. Biol.* **1,** 3–4.
36. Michel, F. & Westhof, E. (1990) *J. Mol. Biol.* **216,** 585–610.
37. Cate, J. H., Gooding, A. R., Podell, E., Zhou, K., Golden, B. L., Kundrot, C. E., Cech, T. R. & Doudna, J. A. (1996) *Science* **273,** 1678–1685.
38. Tanner, M. A. & Cech, T. R. (1997) *Science,* **275,** 847–849.
39. Wang, J.-F., Downs, W. D. & Cech, T. R. (1993) *Science* **260,** 504–508.
40. Celander, D. W. & Cech, T. R. (1991) *Science* **251,** 401–407.
41. von Ahsen, U. & Noller, H. F. (1993) *EMBO J.* **12,** 4747–4754.

Biochemistry **1997**, *36*, 13718−13728

Determinants of Ribose Specificity in RNA Polymerization: Effects of Mn^{2+} and Deoxynucleoside Monophosphate Incorporation into Transcripts†

Ying Huang,‡ Amber Beaudry,§ Jim McSwiggen,§ and Rui Sousa*,‡

Department of Biochemistry, University of Texas Health Science Center, 7703 Floyd Curl Drive, San Antonio, Texas 78284-7760, and Ribozyme Pharmaceuticals Incorporated, 2950 Wilderness Place, Boulder, Colorado 80301

Received July 3, 1997; Revised Manuscript Received August 26, 1997®

ABSTRACT: The catalytic specificity of T7 RNA polymerase (RNAP) for ribonucleoside triphosphates vs deoxynucleoside triphosphates $\{(k_{cat}/K_m)_{rNTP}/(k_{cat}/K_m)_{dNTP}\}$ during transcript elongation is ~80. Mutation of tyrosine 639 to phenylalanine reduces specificity by a factor of ~20 and largely eliminates the K_m difference between rNTPs and dNTPs. The remaining specificity factor of ~4 is k_{cat}-mediated and is nearly eliminated if Mn^{2+} is substituted for Mg^{2+} in the reaction. Mn^{2+} substitution does not significantly affect the K_m difference between rNTPs and dNTPs. Mn^{2+} substitution also enhances the activity of poorly active mutant enzymes carrying nonconservative substitutions in the active site, and its effects are generally consistent with the Mn^{2+}-catalyzed reaction being less restrictive in its requirements for alignment of the reactive groups. In addition to discrimination occurring at the level of nucleoside monophosphate (NMP) incorporation, it is also found that transcripts containing deoxynucleoside monophosphates (dNMPs) are more poorly extended than transcripts of canonical structure, though a severe barrier to transcript extension is seen only when the 3′ region of the transcript is heavily substituted with dNMPs. The barrier to extension of transcripts heavily substituted with dNMPs is reduced for sequences known to be amenable to forming A-like helices and is larger for sequences that resist transformation from B-form DNA·DNA structures. The barrier to extension of dNMP-substituted transcripts is also reduced by solution conditions known to destabilize B-form DNA and to stabilize A-form structures. These observations imply a requirement for a non-B-form, possibly A-like, conformation in the transcript·template hybrid that is disrupted when the transcript is of predominantly deoxyribose structure.

The molecular mechanisms used by nucleic acid polymerases to discriminate substrate structure so as to specify the synthesis of particular products (DNA, RNA, or nucleic acids incorporating noncanonical nucleotides) have, until recently, been obscure. However, over the past few years a number of studies have emerged identifying single amino acid substitutions in the active sites of RNA and DNA polymerases that have dramatic effects on discrimination between ribose 3′ H- or OH-substituted triphosphates (*2*), and between ribose 2′ H- or OH-substitutions (*3*−*5*; see ref *6* for a review).

During transcript elongation T7 RNAP[1] incorporates rNMPs with, on average, ~80-fold higher efficiency than it incorporates dNMPs. The value of ~80 for the $(k_{cat}/K_m)_{rNTP}/(k_{cat}/K_m)_{dNTP}$ term during elongation is accounted for by an average ~20-fold difference in K_m values and an average ~4-fold difference in k_{cat} values. The difference in K_m values is essentially eliminated by mutation of active-site tyrosine 639 to phenylalanine (*5*). The effect of this mutation is remarkably similar to the effect of a Y to F substitution at position 34 in tyrosyl-tRNA synthetase (*7*): in both cases the Y to F substitution has little effect on kinetic parameters for utilization of the canonical substrate (rNTP or tyrosine) but enhances utilization of a noncanonical substrate (dNTP or phenylalanine) that differs from the canonical substrate only in having a hydroxyl group replaced by a hydrogen atom. In both cases the molecular mechanism of the reduced specificity of the mutant enzyme is attributed to loss of a hydrogen bond between the tyrosine hydroxyl and the relevant -OH group on the canonical substrate. That the loss of this H-bonding interaction gives rise to reduced specificity through enhanced utilization of the noncanonical substrate (rather than reduced utilization of the canonical substrate) is attributed to the fact that this interaction makes little *net* contribution to polymerase·substrate association because a water molecule H-bonded to the tyrosine hydroxyl must be displaced to allow catalytically optimal substrate binding.

Previous results have also suggested that, in addition to mechanisms that discriminate at the level of substrate incorporation, extension of transcripts containing dNMPs is less efficient than extension of transcripts of canonical structure (*3*). The determinants of specificity in transcript extension—as opposed to substrate incorporation—have not yet been elucidated. Furthermore, while substitution of Mg^{2+} with Mn^{2+} in the synthesis reaction was found to reduce the substrate specificity of both the wt and Y639F enzymes, the mechanism of the Mn^{2+} effect on substrate specificity and enzyme activity was not addressed in detail.

† Supported by NIH Grant GM52522-01 (to R.S.) and funds from the Texas ARP/ATP.

* To whom correspondence should be addressed: Phone 210-567-3760; Fax 210-567-6595; e-mail sousa@bioc02.uthscsa.edu.

‡ University of Texas Health Science Center.

§ Ribozyme Pharmaceuticals Inc.

® Abstract published in *Advance ACS Abstracts,* October 15, 1997.

[1] Abbreviations: RNAP, RNA polymerase; DNAP, DNA polymerase; dNTP, deoxynucleoside triphosphate; rNTP, ribonucleoside triphosphate; rNMP, ribonucleoside monophosphate; dNMP, deoxynucleoside monophosphate.

S0006-2960(97)01609-7 CCC: $14.00

Here we address these questions. We find that whether a transcript can be efficiently extended is not primarily determined by whether the 3′-NMP is a dNMP or an rNMP but by the level of dNMP substitution in the 3′ region of the transcript. This appears to reflect a requirement for a specific helix conformation in the transcript·template hybrid that is disrupted when the transcript is composed primarily of dNMPs. We also find that the Mn^{2+}-catalyzed reaction is less sensitive to nonconservative active-site mutations and that the reduced substrate specificity of the polymerase in the presence of Mn^{2+} is due largely to an increase in k_{cat} for the noncanonical substrate relative to the canonical substrate. We interpret these observations as indicating that the Mn^{2+}-catalyzed reaction is compatible with a wider range of reactive group geometries than is the Mg^{2+}-catalyzed reaction.

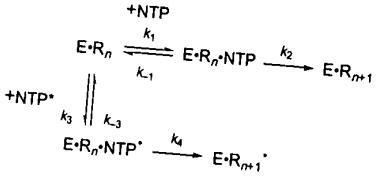
MATERIALS AND METHODS

Mutant and wt enzymes were purified as described (*8*). Activity determinations and NTP K_m determinations were carried out as described (*5*) using supercoiled pT75, which carries a single copy of the T7 ϕ10 promoter (*9*), or homopolymers as templates.

Kinetic parameters for utilization of rGTP and dGTP as elongating substrates for dinucleotide synthesis were also determined as described in ref *5* with reaction conditions modified as follows for particular experiments: (1) Substitution of Mg^{2+} with Mn^{2+} in the reaction was done by replacing the 15 mM $MgCl_2$ in the standard reaction buffer with 15 mM $MnCl_2$ and (2) methanol or ethanol was added to the some reactions to a final percentage (v/v) as specified in the tables and figure captions while maintaining all other reactant and buffer components at constant (moles/volume) concentrations.

Meaning of the rNTP/dNTP Selectivity Term Obtained in the Incorporation Assays. Determination of the catalytic specificity of T7 RNAP for rNTPs vs dNTPs during elongation (as presented in Tables 1 and 8) was carried out as described in ref *5*. In these assays the four unlabeled rNTPs are present at 0.5 mM and a single α-^{32}P-labeled rNTP or dNTP is added to a final concentration of 33 nM. The ratio of the rate of percent incorporation of the radioactive rNTP relative to the radioactive dNTP is then equal to $(k_{cat,rNTP}/K_{m,rNTP})/(k_{cat,dNTP}/K_{m,dNTP})$, as is shown below.

Consider the reaction scheme for an elongation complex that can incorporate either a cold NTP or a radioactively labeled NTP* ($E{\cdot}R_n$ = elongation complex containing transcript of length n):



The rate of incorporation of the labeled NTP is then equal to

$$V_{NTP} = \frac{k_4[E_{tot}]K_{m1}[NTP^*]}{K_{m2}K_{m1} + [NTP^*]K_{m1} + [NTP]K_{m2}}$$

where $K_{m1} = (k_2 + k_{-1})/k_1$ and $K_{m2} = (k_4 + k_{-3})/k_3$. If the cold NTP is present in great excess of the labeled NTP (i.e., in our case [NTP] = ~10 000[NTP*]) and K_{m1} is not much larger than K_{m2} (in our case $K_{m1} \leq K_{m2}$), then the above equation, for *fractional* incorporation, reduces to

$$\frac{V_{NTP^*}}{[NTP^*]} = \frac{k_4[E_{tot}]K_{m1}}{K_{m2}K_{m1} + [NTP]K_{m2}}$$

or

$$\frac{V_{NTP^*}}{[NTP^*]} = \frac{k_4[E_{tot}]K_{m1}/K_{m2}}{K_{m1} + [NTP]}$$

If the labeled NTP is chemically identical to the unlabeled NTP (i.e., both are rNTPs) then obviously $K_{m1} = K_{m2} = K_{m,rNTP}$ and $k_2 = k_4 = k_{cat,rNTP}$. However, if the labeled NTP is a dNTP, then $K_{m1} = K_{m,rNTP} \neq K_{m2} = K_{m,dNTP}$ and $k_2 = k_{cat,rNTP} \neq k_4 = k_{cat,dNTP}$. The ratio of the rate of fractional incorporation of the labeled rNTP relative to the dNTP is then equal to

$$(V_{rNTP^*}/[rNTP^*])/(V_{dNTP^*}/[dNTP^*]) = \{k_{cat,rNTP}[E_{tot}]/(K_{m,rNTP} + [NTP])\}/\{(k_{cat,dNTP}[E_{tot}]K_{m,rNTP}/K_{m,dNTP})/(K_{m,rNTP} + [NTP])\} = (k_{cat,rNTP}/K_{m,rNTP})/(k_{cat,dNTP}/K_{m,dNTP})$$

Kinetic parameters for utilization of rGTP and dGTP as initiating substrates for dinucleotide synthesis were determined using the synthetic hairpin T7 promoter template described in (ref *5*) (HP-GA) in 20 μL reactions containing 15 mM $MgCl_2$, 40 mM Tris, pH 8.0, 5 mM DTT, 1 mM EDTA, with [template] = 1×10^{-7} M, [polymerase] = 1×10^{-6} M, [ATP] = 0.5 mM, and dGTP or rGTP varying from 0.018 to 2.25 mM in serial 2-fold dilutions. [α-^{32}P]ATP was added in trace amounts for labeling. Reactions were initiated by addition of polymerase and were run at 37 °C. Aliquots (5 μL) were taken at 10 and 20 min, mixed with an equal volume of 90% formamide, 0.01% xylene cyanol, and 50 mM EDTA, and electrophoresed on 20% acrylamide/4% bisacrylamide/6 M urea gels. Dinucleotide and unincorporated radioactivity (mononucleotide) bands were quantitated using a Molecular Dynamics phosphorimager. Percent incorporation was translated into the rate of dinucleotide synthesis as a function of [GTP] and rate data were fit by nonlinear least squares (NONLIN) using the Michaelis–Menten equation to derive K_m and k_{cat} values.

Determination of the extension efficiency of transcripts of different ribose/deoxyribose composition was carried out at 37 °C for 10 min in 25 μL reaction volumes using pT75 as a template at 10^{-7} M and wt or Y639F polymerase at 10^{-8} M in 15 mM $MgCl_2$, 40 mM Tris, pH 8.0, 5 mM DTT, and 1 mM EDTA. Varying combinations of rNTPs, dNTPs, and rGMP were used in different reactions as indicated in the tables and figures. rNTPs were always present at 0.5 mM, and dNTPs and rGMP were added to final concentrations of 1 mM. Reactions contained either [α-^{32}P]GTP or dGTP for labeling, depending on whether cold rGTP or dGTP was being used in the reaction. Reactions were terminated by addition of an equal volume of 90% formamide, 0.01% xylene cyanol, and 50 mM EDTA and electrophoresed on 20% acrylamide/4% bisacrylamide/6 M urea gels, which were quantitated on a Molecular Dynamics phosphorimager.

Table 1: Relative Activity and rNTP/dNTP Specificity of wt and Y639 Mutant Polymerases in Mg^{2+} or Mn^{2+} Catalyzed Reactions[a]

RNAP	activity	A	C	G	U	average
			Mg^{2+}			
wt	200	121 ± 33	89 ± 13	60 ± 11	34 ± 8.5	76
Y639F	200 ± 32	5.0 ± 1.8	7.5 ± 2.0	2.0 ± 0.41	1.8 ± 0.39	4.1
Y639M	104 ± 15	8.8 ± 1.5	8.5 ± 2.0	2.3 ± 0.41	2.5 ± 0.84	5.5
Y639L	86 ± 12	8.2 ± 1.1	21 ± 4.7	12 ± 3.2	4.0 ± 0.97	11
Y639C	15 ± 5.6	14 ± 2.3	14 ± 5.3	9.2 ± 2.3	6.6 ± 2.4	11
Y639V	8.6 ± 4.6	22 ± 6.6	21 ± 3.9	26 ± 5.0	11 ± 1.7	19
Y639T	2.6 ± 0.86	8.8 ± 1.7	7.1 ± 2.4	4.9 ± 1.0	5.0 ± 9.8	6.5
Y639Q	2.0 ± 0.34	5.3 ± 1.5	6.8 ± 1.8	3.5 ± 1.6	2.4 ± 1.4	4.5
Y639H	0.74 ± 0.08	nd	nd	nd	nd	nd
S641A	186 ± 26	125 ± 28	77 ± 10	59 ± 7.4	30 ± 9.0	73
Y639F/S641A	178 ± 22	12 ± 0.85	10 ± 4.2	4.1 ± 1.6	2.7 ± 0.57	7.3
			Mn^{2+}			
wt	21 ± 1.9	40 ± 14	18 ± 2.5	14 ± 5.6	9.6 ± 3.7	20
Y639F	12 ± 1.2	1.9 ± 0.50	1.7 ± 0.50	1.2 ± 0.34	2.0 ± 0.6	1.7
Y639M	11 ± 0.96	3.3 ± 0.25	1.7 ± 0.15	1.2 ± 0.50	2.2 ± 0.65	2.1
Y639L	13 ± 1.4	3.1 ± 0.15	4.2 ± 0.50	2.6 ± 0.25	2.5 ± 0.55	3.1
Y639C	16 ± 4.0	1.0 ± 0.05	1.0 ± 0.50	0.97 ± 0.14	1.8 ± 0.30	1.2
Y639V	14 ± 2.4	32 ± 8.1	14 ± 4.6	7.1 ± 1.2	4.8 ± 0.60	14
Y639T	5.2 ± 1.0	25 ± 8.9	13 ± 3.5	4.3 ± 1.4	6.1 ± 0.25	12
Y639Q	3.4 ± 0.34	5.1 ± 1.5	1.9 ± 0.30	1.5 ± 0.35	1.4 ± 0.34	2.3
Y639H	1.0 ± 0.19	nd	nd	nd	nd	nd

[a] A, C, G, and U indicate the $(k_{cat}/K_m)_{rNTP}/(k_{cat}/K_m)_{dNTP}$ values for the NTP (see Materials and Methods) with the corresponding base. Average indicates the average rNTP/dNTP catalytic specificity factor for all four NTPs. Data for reactions with Mg^{2+} have been presented previously (*5*) and are presented for comparison. Values given are ±SE.

Table 2: Kinetic Parameters for wt T7 RNAP Utilization of rUTP and dUTP during Elongation in Mg^{2+} and Mn^{2+} Reactions[a]

cofactor	$K_{m,rU}$ (mM)	$K_{m,dU}$ (mM)	$\frac{(k_{cat}/K_m)_{rU}}{(k_{cat}/K_m)_{dU}}$	$\frac{\{(k_{cat}/K_m)_{rU}/(k_{cat}/K_m)_{dU}\}_{Mg}}{\{(k_{cat}/K_m)_{rU}/(k_{cat}/K_m)_{dU}\}_{Mn}}$	$\frac{(K_{m,dU}/K_{m,rU})_{Mg}}{(K_{m,dU}/K_{m,rU})_{Mn}}$	$\frac{(k_{catrU}/k_{catdU})_{Mg}}{(k_{catrU}/k_{catdU})_{Mn}}$
Mg^{2+}	0.041 ± 0.018	1.7 ± 0.33	35 ± 9.5	3.6	1.2	3.0
Mn^{2+}	0.036 ± 0.028	1.2 ± 0.077	9.6 ± 3.7			

[a] Values given are ±SE for $n = 3$.

Poly(dG) and poly(rG) ladders generated in reactions in which dGTP or rGTP was the only NTP present were used as markers for assigning the length and sequence of different transcripts. Percent extension for a transcript of length n was determined by dividing the molar amount of all transcripts of length $>n$ by the molar amount of all transcripts $\geq n$.

RESULTS

Effect of Substituting Mn^{2+} for Mg^{2+} on Specificity and Activity. Table 1 presents values for the effect of Mn^{2+} substitution on the relative activity and rNTP/dNTP specificity of polymerases carrying different substitutions at position 639. Substituting Mn^{2+} for Mg^{2+} in the reaction reduces wt activity by ~10-fold on a plasmid template. Nonconservative substitutions at position 639 were previously shown to markedly reduce activity (*5*), but with Mn^{2+} the effects of nonconservative substitutions are ameliorated. Thus Y639T, for example, is ~75-fold less active than the wt enzyme in Mg^{2+} buffer but is only ~4-fold less active than wt in Mn^{2+} buffer. Mn^{2+} substitution also affects rNTP/dNTP specificity $\{(k_{cat}/K_m)_{dNTP}/(k_{cat}/K_m)_{rNTP}\}$, reducing it by an average factor of ~4 for the wt enzyme and by ~3-fold for Y639F.

Effect of Mn^{2+} Substitution on Kinetic Parameters for dNTPs and rNTPs. The reduction in rNTP/dNTP selectivity obtained with Mn^{2+} could be due to effects on the relative K_m and/or k_{cat} values for dNTPs vs rNTPs. To determine which parameter was being affected by Mn^{2+} we measured NTP K_m values on a supercoiled plasmid template in Mg^{2+} and Mn^{2+} buffer (Table 2). Measurement of kinetic parameters on such templates is complicated by the occurrence of phases of transcription (initiation, abortive cycling, elongation) with distinct processivities and different apparent NTP K_m values. Previously we found that, with supercoiled templates in Mg^{2+} buffer, the elongation phase of the reaction dominated the kinetics so that NTP K_m values for elongation could be determined (*5*). However, when we examined transcript patterns in reactions with the wt enzyme and with one rNTP substituted with a dNTP in Mn^{2+} buffer, it was apparent that the reduced activity of the wt enzyme under these conditions greatly increased the level of abortive cycling (not shown) and the assumption that the reaction kinetics would be dominated by the elongation phase of the transcription reaction no longer appeared valid. We therefore limited our measurement of elongation phase K_m values to rUTP and dUTP because, with the canonical T7 promoters used in this study, these nucleotides are incorporated only during the elongation phase of transcription. While K_m values can be determined in this way, k_{cat} values for dUMP incorporation cannot be determined directly in these assays because it is not possible to separate the contributions to V_{max} of a reduced incorporation rate for dUMP and a reduced rate of extension of dUMP substituted transcripts. However, the ratio of $k_{cat,rU}/k_{cat,dU}$ can be determined from the catalytic specificity $(k_{cat}/K_m)_{rU}/(k_{cat}/K_m)_{dU}$ determined previously (Table 1) and from the ratio of $K_{m,dU}/K_{m,rU}$. We found that Mn^{2+} substitution had little or no effect on the ratio of the K_m values

Table 3: Kinetic Parameters for Utilization of rGTP and dGTP as the Elongating Substrate during Dinucleotide Synthesis in Mg^{2+} and Mn^{2+} Reactions[a]

pol (cofactor)	$K_{m,rG}$ (μM)	$K_{m,dG}$ (μM)	$k_{cat,rG}$	$k_{cat,dG}$	$\frac{(k_{cat}/K_m)_{rG}}{(k_{cat}/K_m)_{dG}}$	$\frac{\{(k_{cat}/K_m)_{rG}/(k_{cat}/K_m)_{dG}\}_{Mg}}{\{(k_{cat}/K_m)_{rG}/(k_{cat}/K_m)_{dG}\}_{Mg}}$	$\frac{(k_{cat,rG}/k_{cat,dG})_{Mg}}{(k_{cat,rG}/k_{cat,dG})_{Mn}}$	$\frac{(K_{m,dG}/K_{m,rG})_{Mg}}{(K_{m,dG}/K_{m,rG})_{Mn}}$
wt (Mg)	17.5 ± 2.5	850 ± 76	2.55 ± 0.15	0.018 ± 0.001	6.9×10^3			
wt (Mn)	10.3 ± 3.3	387 ± 23	1.2 ± 0.15	0.17 ± 0.034	2.7×10^2	26	20	1.3
Y639F (Mg)	10.4 ± 0.65	15 ± 0.70	1.8 ± 0.1	0.054 ± 0.005	50			
Y639F(Mn)	15.3 ± 2.7	25 ± 7.0	1.1 ± 0.25	0.75 ± 0.16	2.4	21	23	0.9

[a] Data for reactions in Mg^{2+} buffer have been presented previously (*5*) and are presented here for comparison. Values given are ±SE for $n = 3$.

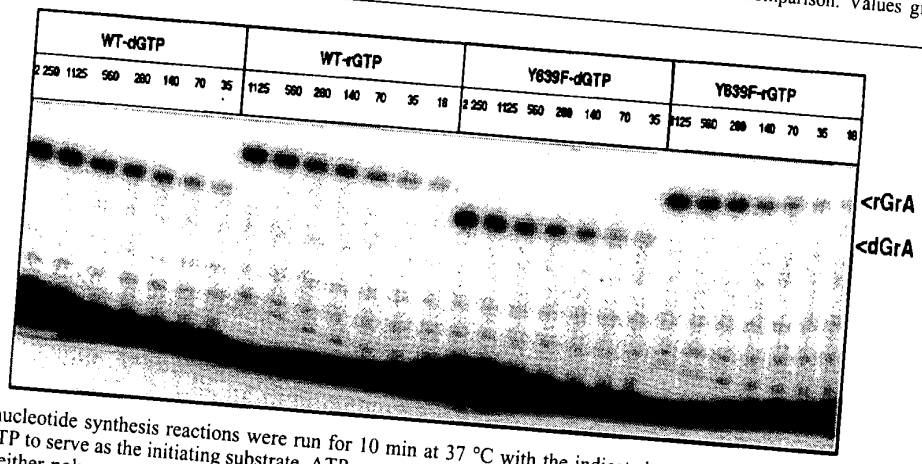


FIGURE 1: Dinucleotide synthesis reactions were run for 10 min at 37 °C with the indicated polymerases and micromolar concentrations of rGTP or dGTP to serve as the initiating substrate. ATP was present at 0.5 mM in all reactions and [α-^{32}P]ATP was used in trace amounts for labeling. Neither polymerase shows strong discrimination against utilization of dGTP as an initiating nucleotide.

for rUTP and dUTP, which allows us to attribute the reduced substrate specificity in Mn^{2+} buffer largely to an effect on the relative k_{cat} values for these substrates (Table 2). Since both ATP and GTP are incorporated during abortive cycling on all canonical T7 promoters, K_m values for purines could not be evaluated with this assay. We therefore used dinucleotide synthesis to evaluate the effects of Mn^{2+} substitution on the kinetic parameters for utilization of rGTP and dGTP as elongating substrates (Table 3). This assay also allows apparent k_{cat} for incorporation of dGMP to be determined directly without having to consider the effect of dGMP incorporation on subsequent extension. Dinucleotide synthesis is characterized by a much higher level of rGTP/dGTP specificity than the elongation reaction. This is primarily due to a decrease in k_{cat} for incorporation of dGTP relative to rGTP (Table 3) in dinucleotide synthesis vs elongation. The effects of the Y639F mutation and Mn^{2+} substitution are unambiguously distinct. The Y639F mutation eliminates the 40−50-fold K_m difference between rGTP and dGTP but affects the ~140-fold difference in k_{cat} between these two substrates by only a factor of ~4. Mn^{2+} substitution, however, reduces the ratio of the k_{cat} values of rGTP and dGTP by more than 20-fold but has no significant effect on their relative K_m values.

Kinetic Parameters for Initiation with a dNTP vs an rNTP. In order to characterize kinetic parameters for utilization of a dNTP at the initiating (or 3′-NMP) substrate site we measured the rates of GA dinucleotide synthesis on a self-complementary hairpin−promoter template in the presence of saturating concentrations of the elongating substrate (ATP) and varying concentrations of GTP or dGTP. The data in Table 3 reveal that, in Mg^{2+} buffer, wt enzyme utilization of dGTP as an *elongating* substrate in a dinucleotide synthesis reaction is characterized by a ~48-fold higher K_m and a ~140-fold lower k_{cat} than utilization of rGTP. However, while the use of a dNTP in the elongating substrate site is strongly discriminated against by the wt enzyme, the utilization of a dNTP as an initiating substrate is only modestly disfavored (Figure 1): the wt enzyme shows only a 3-fold higher K_m and a virtually identical k_{cat} when using dGTP vs rGTP for initiation (Table 4). The Y639F and S641A mutations and the Y639F/S641A double mutation do not markedly affect these kinetic parameters (Table 4). The S641A mutation was tested because of a previous report that this substitution has a dramatic effect on the substrate/product specificity of T7 RNAP (*10*). Previously, we showed that, contrary to this report, the S641A substitution does not affect the specificity of incorporation of the elongating substrate (*5*). As shown in Table 4 it also appears that the S641A substitution does not affect specificity for the initiating substrate.

Table 4: Kinetic Parameters for the Initiating Nucleotide in rGrA or dGrA Synthesis[a]

	wt	Y639F	S641A	Y639F/S641A
$K_{m,rGTP}$ (mM)	0.25 ± 0.04	0.32 ± 0.02	0.21 ± 0.04	0.22 ± 0.090
$K_{m,dGTP}$ (mM)	0.88 ± 0.11	0.75 ± 0.18	1.4 ± 0.40	1.1 ± 0.12
$k_{cat,rGTP}$ (s^{-1})	0.26 ± 0.07	0.28 ± 0.08	0.38 ± 0.06	0.25 ± 0.03
$k_{cat,dGTP}$ (s^{-1})	0.22 ± 0.02	0.34 ± 0.05	0.32 ± 0.07	0.10 ± 0.02

[a] Values are ±SE for $n = 3$. Kinetic parameters were determined from time points taken at 10 and 20 min, over which incorporation rates were found to be linear.

Effects on the Extension Reaction of Having dNMPs Incorporated in the Transcript. Results from the dinucleotide synthesis reaction revealed that the polymerase discriminates only weakly against using a dNTP as an initiating substrate. Since the initiating substrate occupies the same site as the

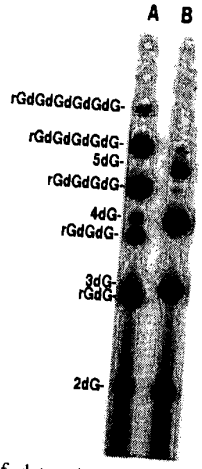


FIGURE 2: Method of determination of percent extension of transcripts of varying rNMP/dNMP composition. Lane A shows reactions run with Y639F and with 0.5 mM dGTP, 0.5 mM rGMP, and [α-^{32}P]dGTP; lane B shows similar reactions run without rGMP. The structure of the products is indicated. The reduced charge on transcripts that are 5′-terminated with rGMP causes them to migrate more slowly than 5′-dGTP-terminated transcripts of similar size. Percent extension of a transcript of length n is determined by dividing the molar amount of all transcripts > n by all transcripts $\geq n$ [pure oligo(dG) transcripts obtained in lane A are not included in the calculation to determine efficiency of extension of 5′-rGMP-terminated transcripts]. Qualitatively, the levels of extension of the 2dG (lane B) or rGdG (lane A) transcripts appear similar. However, the rGdGdG transcript is clearly extended more efficiently than the 3dG transcript, and the rGdGdGdG transcript is also seen to be extended more efficiently than the 4dG transcript. However, the rGdGdGdGdG transcript is poorly extended, indicating that the single rG in this position no longer enhances transcript extension, possibly because it no longer base-pairs with template (wobble base-pairing with the T at −1 could extend the oligo G hybrid to 4 base pairs).

3′-NMP of the transcript during catalysis (*11*), this could imply that the polymerase does not discriminate against a transcript with a 3′-dNMP. However, it is possible that in the context of a more extended helical structure the presence of a 3′-dNMP in the transcript could be more detrimental to the extension reaction. To evaluate this, we measured the efficiency of extension of transcripts of varying rNMP/dNMP composition generated in situ during initial transcription (see Figure 2 for an illustrative example of this method). These data are presented in Table 5, from which it is clear that, with the wt enzyme, extension reactions involving incorporation of a dNMP are much less efficient than corresponding reactions (identical substrate base and transcript sequence) involving incorporation of an rNMP. It is also clear that the Y639F mutation increases the efficiency of extension in reactions with dNTP substrates. However, in addition to the wt enzyme displaying specificity at the level of substrate incorporation, it is also seen that both the wt and mutant enzymes display specificity for transcript structure in that they extend transcripts containing dNMPs less efficiently than transcripts of identical sequence that are composed entirely of rNMPs. It is found, though, that low to moderate levels of dNMP incorporation in the transcripts give rise to only modest decreases in the efficiency with which those transcripts are extended, nor does the presence of a 3′-dNMP generate a particularly strong barrier to transcript extension. For example, rGrGrGdA transcripts are extended by either the wt or Y639F enzymes with rG, with efficiency similar

Table 5: Percent Extension of Transcripts of Varying rNMP/dNMP Composition with rNTPs or dNTPs[a]

extension reaction	wt	Y639F
rGrG + rG	63−65	57−63
rG**dG** + **dG**	22−25	56−61
dGdG + **dG**	13−27	60−64
rGrGrG + rA	96−97	86−90
rGrGrG + **dA**	1−9	41−57
rG**dGdG** + rA	85−95	50−54
dGdGdG + rA	nd	10−16
rGrGrG + rG	77−81	68−74
rG**dGdG** + **dG**	61−75	88−90
dGdGdG + **dG**	6−14	49−57
rGrGrGrA + rG	93−95	69−76
rGrGrG**dA** + rG	92−98	66−80
rGrGrGrG + rG	69−69	58−61
rG**dGdG**rA + **dG**	3−5	31−36
rG**dGdGdG** + **dG**	2−6	20−30
dGdGdGdG + **dG**	nd	8−13
rGrGrGrArG + rA	97−99	92−94
rGrGrG**dA**rG + **dA**	50−70	82−85
rGrGrGrGrG + rG	73−73	61−64
rG**dGdG**r**AdG** + rA	nd	60−69
rG**dGdGdGdG** + **dG**	nd	4−4
rGrGrGrArGrA + rC	96−99	95−96
rGrGrGrArGrA + **dC**	20−31	92−94
rGrGrG**dA**rG**dA** + rC	77−78	65−72
rGrGrG**dA**rG**dA** + **dC**	11−14	40−41
rG**dGdG**r**AdG**rA + rC	nd	69−72
rGrGrGrArGrArC + rC	96−99	96−98
rGrGrGrArGrA**dC** + **dC**	35−53	88−94
rGrGrG**dA**rG**dA**rC + rC	95−95	91−93
rGrGrG**dA**rG**dAdC** + **dC**	nd	32−34
rG**dGdG**r**AdG**rArC + rC	nd	75−84
rGrGrGrArGrArCrC + rG	90−98	82−85
rGrGrGrArGrA**dCdC** + rG	81−84	71−76
rGrGrG**dA**rG**dA**rCrC + rG	87−94	77−83
rGrGrG**dA**rG**dAdCdC** + rG	nd	15−22

[a] Values given are ranges from two experiments. Since transcripts are generated in situ, during initial transcription, transcripts with multiple deoxyribose substitutions are obtained at low levels with the wt enzyme. In many cases the level of a particular transcript (e.g., dGdGdGdG) generated by the wt enzyme is too low to allow quantitation of its percent extension (nd).

to extension of rGrGrGrA transcripts. Strong decreases in the efficiency of extension that are attributable to transcript structure occur only when (1) the transcript is at least 3 bases in length and (2) it is heavily substituted with dNMPs in the 3−5 bases nearest the 3′-end of the transcript. It is therefore seen that transcripts 3 and 4 bases in length that are 5′-terminated with rGMP but contain dGMP at the internal or 3′-positions are extended with markedly greater efficiency than transcripts of otherwise identical structure but that carry a dGTP at the 5′-position (Table 5 and Figure 2). A similar effect is observed if the relative extension efficiencies of the rGrGrGrArGrAdCdC and rGrGrGdArGdAdCdC transcripts are compared. Thus dNMP substitution at positions up to 2−4 bases away from the 3′-base of the transcript can affect the efficiency with which the transcript is extended. Together with the results from the dinucleotide synthesis reactions, these observations imply (1) that the Y639F mutation enhances dNMP incorporation but does not affect extension of dNMP-substituted transcripts, (2) the presence of a 3′ dNMP or a moderate level of dNMP substitution in the 3′ segment of the transcript only modestly reduces extension efficiency, and (3) a high level of dNMP substitu-

Table 6: Relative Activity of wt and Y639F Polymerases with dNTPs and rNTPs on Different Templates[a]

	poly(dI)·poly(dC)			poly(dA)·poly(dT)			pT75			
	rGTP	dGTP	dGTP + rGMP	rATP	dATP	dATP + rAMP	rGTP	rITP	dGTP	dITP
wt	160	2.0 (0.9−2.5)	4.0 (3.2−4.4)	32 (28−46)	<0.1	<0.1	200	233 (217−250)	1.1 (1.0−1.2)	<0.1
Y639F	160	156 (108−200)	172 (132−232)	10 (8−12)	<0.1	<0.1	200	210 (194−227)	52 (44−60)	<1.1

[a] The templates and NTPs used are as specified in the table. For the pT75 reaction, rATP, rCTP, and rUTP were also present at 0.5 mM. An activity value of 200 in this and all following tables corresponds to incorporation of 52 μmol of NTP/min in reactions with 10^{-7} M pT75 or 0.1 mg/mL homopolymer and 10^{-8} M polymerase. Data for the poly(dI)·poly(dC) template were published previously (*3*) and are presented here for comparison. Values given are mean and ranges from three (homopolymers) or two (pT75) experiments.

tion in the 3−5 bases nearest the 3′-end of the transcript strongly reduces extension efficiency.

One interpretation of the preceding observations is that the conformation of the transcript·template duplex in the active site is an important determinant of the efficiency of the extension reaction. For example, the canonical transcript·template conformation could be an A-form RNA·DNA hybrid and the extension reaction might be compromised if this duplex became B-form as a consequence of extensive dNMP substitution in the transcript. In addition to being influenced by ribose structure, duplex nucleic acid conformation is strongly influenced by water activity, sequence, and base composition (*12*−*16*). In particular, poly(dA)·poly(dT) forms a B-type structure and resists forming an A-like structure even at a relative humidity of 0%. Alternating purine−pyrimidine sequences tend to stabilize B-form DNA (*17*), while guanosine-rich DNA is particularly amenable to forming A-type structures. Poly(dG)·poly(dC) would then be expected to be especially amenable to forming an A-type structure, and in fact this is the only DNA duplex known to be A-form in solution (*18*). We therefore measured the activity of the wt and Y639F enzymes on homopolymeric templates (T7 RNAP efficiently synthesizes polypurine transcripts from polypyrimidine templates in a promoter-independent manner). As shown in Table 6, the Y639F enzyme synthesizes either poly(rG) or poly(dG) with poly(dI)·poly(dC) as a template with essentially indistinguishable activity under these assay conditions. Even the wt enzyme displays significant activity in poly(dG) synthesis on the poly(dC) template. This is remarkable given that Y639F is ≥600-fold less active in synthesis of pure DNA vs pure RNA on a complex sequence template (Table 7). This cannot be attributed simply to the fact that poly(dC) is a homopolymer template since replacement of rATP with dATP in synthesis reactions with the poly(dA)·poly(dT) template completely abolishes the activity of both the wt and Y639F enzymes. This was further evaluated by measuring the relative activity of the wt and Y639F polymerases on supercoiled pT75 in reactions in which rGTP was replaced by rITP, dGTP, or dITP (Table 6). dI·dC base pairs are like dA·dT base pairs in that they strongly favor assumption of B-form conformations and resist transformation to A-form (*14*−*16*). Activity in reactions with rITP was indistinguishable from that in reactions with rGTP, indicating that there was no strong discrimination against incorporation of the inosine base. However, the activity of the Y639F enzyme was reduced by more than 40-fold when dITP replaced dGTP in the reaction, suggesting that the stronger tendency of dI·dC base pairs to form B-type helices may have significantly depressed transcription over the G·C-rich initially transcribed sequence of this promoter.

Solution Conditions That Enhance Extension of Transcripts with Multiple dNMP Substitutions. In light of these results

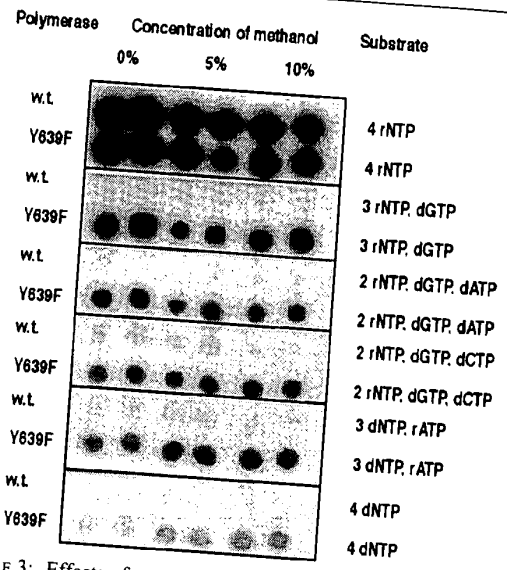


FIGURE 3: Effects of methanol addition on wt and Y639F activity in reactions with different combinations of rNTPs and dNTPs. Transcription reactions carried out with the indicated NTPs, polymerases, and methanol concentrations and supercoiled pT75 as template were carried out for 40 min at 37 °C, spotted onto DE81 filter paper, and washed with 0.5 M Na_2HPO_4 to remove unincorporated radioactivity (either [α-^{32}P]rGTP or dGTP). Methanol addition markedly enhances activity in the 3 dNTP and 4 dNTP reactions.

and previous observations that alcohols—through their dehydrating and water-structure perturbing effects—affect helix conformation by destabilizing B-form structures (and thereby favor A-form helices) (*1*, *12*, *13*, *19*) we decided to test the effects of methanol or ethanol addition to transcription reactions with varying combinations of rNTPs and dNTPs (Figure 3, Table 7). Table 7 reveals that addition of 15% methanol to reactions with four rNTPs reduces the activity of both the wt or Y639F polymerases by 4−5-fold. Substitution of rGTP by dGTP in the reaction with the wt enzyme reduces activity by 50−200-fold (depending on the methanol concentration), though activity with the Y639F mutant is reduced only 1−4-fold. Addition of methanol to reactions in which one or two rNTPs were replaced with dNTPs does not significantly increase absolute levels of Y639F activity; however, addition of methanol to such reactions does increase activity relative to reactions carried out with four rNTPs at the same methanol concentration. These results suggest that methanol has at least two effects on the reaction: a general decrease in enzyme activity with increasing methanol concentrations (ethanol was found to depress activity to an even greater degree and was therefore less useful in this regard) is coupled with decreasing sensitivity to replacement of rNTPs with dNTPs in the reaction. The former effect is clearly evidenced in 4 rNTP

Table 7: Effect of Methanol and Ethanol on Relative Activity in Reactions with Different Combinations of rNTPs and dNTPs[a]

NTPs (alcohol)	wt alcohol (v/v)					Y639F alcohol (v/v)				
	0%	5%	10%	15%	20%	0%	5%	10%	15%	20%
4 rNTPs (MeOH)	200	167 ± 27	96 ± 15	54 ± 5	42 ± 6	200	138 ± 30	81 ± 10	41 ± 3	25 ± 3
4 rNTPs (EtOH)	200	82 ± 7	52 ± 5	13 ± 2	8.8 ± 1.6	200	80 ± 12	39 ± 4	16 ± 3	8.0 ± 1.2
3 rNTPs, dGTP (MeOH)	1.1 ± 0.1	1.3 ± 0.1	1.2 ± 0.2	1.4 ± 0.3	1.2 ± 0.2	52 ± 8	54 ± 3	49 ± 3	44 ± 4	30 ± 3
3 rNTPs, dGTP (EtOH)	1.1 ± 0.1	1.5 ± 0.2	1.6 ± 0.2	1.0 ± 0.2	0.42 ± 0.1	52 ± 8	49 ± 7	26 ± 3	14 ± 3	4.5 ± 1.2
2 rNTPs,dA,dG (MeOH)	0.45 ± 0.1	0.36 ± 0.1	0.48 ± 0.1	0.8 ± 0.2	0.8 ± 0.2	14 ± 2	18 ± 3	19 ± 2	21 ± 2	20 ± 3
2 rNTPs,dT,dG (MeOH)	0.33 ± 0.1	0.26 ± 0.1	0.46 ± 0.1	<0.4	<0.4	14 ± 3	16 ± 4	18 ± 5	24 ± 3	22 ± 4
3 dNTPs, rATP (MeOH)	<0.3	<0.3	<0.3	<0.3	<0.3	3.3 ± 1	6.0 ± 0.8	7.1 ± 0.7	8.8 ± 0.8	8.8 ± 1
4 dNTPs (MeOH)	<0.1	<0.1	<0.1	<0.1	<0.1	0.33 ± 0.1	1.1 ± 0.4	2.5 ± 0.4	2.6 ± 0.5	2.4 ± 0.4

[a] Reaction time points were taken at 2.5, 5, 10, 20, and 40 min. Incorporation was linear under all conditions for the first four time points, so only data for the 2.5−20 min time points were used to determine relative activity. Values given are ±SE for $n = 3$.

Table 8: Effect of Methanol on $(k_{cat}/K_m)_{rGTP}/(k_{cat}/K_m)_{dGTP}$[a]

	methanol (v/v)				
	0%	5%	10%	15%	20%
wt	75−83	75−93	77−78	79−98	84−86
Y639F	2.9−3.2	2.5−3.4	2.5−3.2	2.8−3.1	3.1−3.4

[a] Values given are ranges from two experiments.

reactions, while in the 1 dNTP or 2 dNTP reactions the two effects appear balanced and little net effect on absolute activity is obtained due to methanol addition. However, in the 3 dNTP or 4 dNTP reactions the reduced sensitivity to dNTPs dominates, and addition of 15% methanol to the reaction with Y639F leads to substantial (~8-fold) increases in absolute activity. Though this is a substantial net increase, the degree to which methanol addition reduces sensitivity to multiple dNTP replacement in the transcription reaction is actually much greater. For example, in the absence of methanol, replacement of all four rNTPs with dNTPs reduces Y639F activity by ≥600 fold. However, in the presence of 20% methanol activity is only ~10-fold lower in the 4 dNTP reaction than in the 4 rNTP reaction, representing a net 60-fold decrease in the degree to which the reaction is sensitive to replacement of all four rNTPs with dNTPs.

Effect of Methanol on Specificity during Incorporation. Reactions in which one or more rNTPs are substituted with dNTPs may be limited at either the dNMP incorporation or dNMP-substituted transcript extension steps. Since the barrier to Y639F synthesis of transcripts multiply substituted with dNMPs appears to be largely due to transcript structure rather than substrate structure, and since methanol addition had its largest effects in the 3 or 4 dNTP reactions, it appeared likely that methanol enhanced the extension of dNMP-substituted transcripts rather than incorporation of dNMPs. To test this, we evaluated the effect of methanol addition on the relative rates of radiolabeled rGTP or dGTP incorporation in reactions in which unlabeled rGTP was present in large excess. Under such conditions the rate of dGTP incorporation is not limited by effects on transcript structure since the fractional substitution of dGMP into the transcripts is less than ~0.01%. We found that methanol addition had little effect on the rGTP/dGTP specificity ratio in such an assay (Table 8), implying that the methanol enhances extension of dNMP-substituted transcripts rather than incorporation of dNMPs. This was also supported by examination of the transcript patterns obtained with the

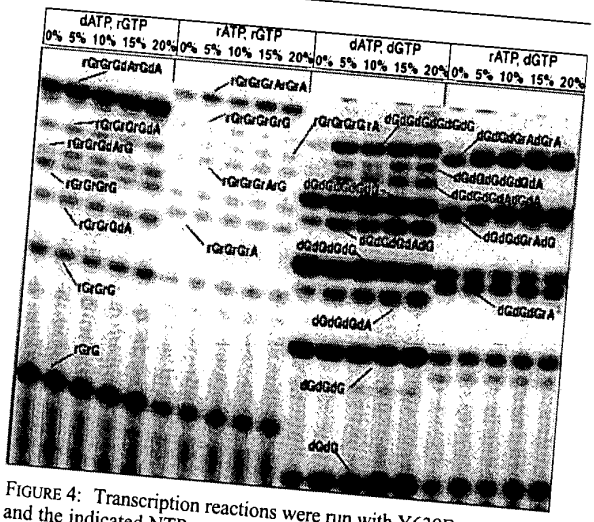


FIGURE 4: Transcription reactions were run with Y639F polymerase and the indicated NTPs on a partially single-stranded synthetic T7 promoter template composed of an 18-base nontemplate strand corresponding to the sequence of the consensus class III T7 promoter from −23 to −5 annealed to a 33-base template strand corresponding to the T7 ϕ10 promoter and ITS (initially transcribed region; from +1 to +6 GGGAGA). [Template] = 10^{-7} M; [pol] = 10^{-6} M; the reaction was run for 15 min at 37 °C before being electrophoresed on 20% acrylamide/4% bisacrylamide/6 M urea gels and visualized by phosphorimaging. Sequence assignments were made by comparison to poly(dG) and poly(rG) oligomers obtained in reactions in which only dGTP or rGTP was present and by comparison of migration patterns of transcripts obtained with the indicated NTPs. The relative migration rates of short transcripts of identical length but varying sequence and ribose composition follows the rule, in order of increasing retardation: dA < rA < dG < rG.

Y639F enzyme in the presence of varying concentrations of methanol and the following combination of NTPs: rATP + rGTP, dATP + rGTP, rATP + dGTP, and dATP + dGTP (Figure 4). The steady-state rate of transcript production in such assays is limited by the rate of transcript release (*20*), and both the incorporation of dNMPs into the transcript and addition of methanol appear to destabilize the complex and therefore lead to net increases in the total amount of transcript obtained. However, in evaluating the effects of methanol on transcript extension or nucleotide incorporation, we focus on the changes in the transcript patterns (processivity) rather than the total transcript levels. The transcript patterns obtained on this promoter are complex: in addition to transcripts corresponding to the sequence of the +1 to +6

region of the promoter (GGGAGA), products with larger numbers of G residues are also obtained due to slippage of the transcript on the initial run of 3 Gs. The G-slippage reaction has invariably been found to be more robust than heterogeneous sequence synthesis: thus, whenever the enzyme is mutated so that it is significantly less active, it has been found that poly-G ladder synthesis predominates during initiation (*21*−*23*). This also occurs when the reaction is less efficient due to noncanonical substrate or transcript structure (*3*); for example, in the 0% methanol, dATP + dGTP reaction in Figure 4, it can be seen that most of the dGdGdG trimer is extended to form larger poly(dG) oligomers rather than the heterogeneous sequence oligomers. Addition of methanol does not affect the transcript pattern in the rATP + rGTP reaction. Nor is there much of an effect on the transcript pattern in the dATP + rGTP reaction. However, in the dGTP + rATP reaction methanol addition has a clear effect: there is an increase in the efficiency of extension of the dGdGdG trimer with rA at the expense of extension to form longer poly(dG) oligomers so that the amount of the 4-dG product decreases while the levels of the longer heterogeneous sequence transcripts increase. In the dATP + dGTP reaction the effect of methanol is even greater: without methanol there is very little production of heterogeneous sequence transcripts and there is little extension of the poly(dG) oligomers beyond 5 bases in length, while addition of increasing concentrations of methanol progressively increases both the levels of heterogeneous sequence transcripts and the efficiency with which the poly(dG) oligomers are extended.

DISCUSSION

Effects of Mn^{2+} on Specificity and Enzyme Activity. Use of Mn^{2+} as a catalytic cofactor has generally been found to cause a reduction in the substrate specificity or fidelity of enzymes that normally use Mg^{2+} to catalyze a phosphoryl transfer reaction. For T7 RNAP we find that the reduced discrimination for incorporation of a 2′-deoxypyrimidine during elongation or incorporation of a 2′-deoxypurine during formation of the first phosphodiester bond is mediated by a reduction in the ratio of the k_{cat} values for rNTP and dNTP substrates, as has also been found for the effect of Mn^{2+} on deoxy/dideoxy discrimination in DNAP I (*24*) and Taq DNAP (*25*). This is a rather critical result in light of our previous conclusion that the K_m difference between these substrates is due to the Y639-OH·ribose 2′-OH interaction. It would have been difficult to understand how Mn^{2+} substitution could influence this interaction. The molecular mechanisms and effects of Mn^{2+} substitution and the Y639F mutation should therefore be complementary and distinct, and this is what is found (Tables 1−3).

Substitution of Mg^{2+} with Mn^{2+} in polymerization reactions leads to a remarkably wide range of effects with many polymerases. In addition to increasing the relative levels of base misincorporation (*26*−*30*), Mn^{2+} increases the relative utilization of substrates modified at different positions on the base, ribose, or phosphate groups (*31*, *32*) as well as relative activity on templates of noncanonical structure (*33*, *34*) or even in the absence of a templating base (*35*). Structural alterations of the polymerase active site also appear to perturb the Mn^{2+}-catalyzed reaction to a lesser degree than the Mg^{2+}-catalyzed reaction. Thus, while substitution of Mn^{2+} for Mg^{2+} markedly reduced the activity of the wt or Y639F enzymes, it either enhanced the activity of the poorly active mutant enzymes bearing nonconservative substitutions at position 639 or it reduced the activity of these mutants to only a small degree (Table 1). Similar observations have been made in other systems (*34*−*37*). Such a broad spectrum of effects is effectively summarized by the statement that the Mn^{2+}-catalyzed reaction is compatible with a wider range of reactive group geometries than is the Mg^{2+}-catalyzed reaction, so that it is less sensitive to changes in reactive group geometries due either to noncanonical substrate structure or to nonconservative active-site mutations. It is as yet unclear what properties of Mn^{2+} relative to Mg^{2+} [greater softness of the Mn^{2+} ion (*38*), faster rates of hydration/dehydration, tighter ligand binding (*39*)] are responsible for its distinctive catalytic properties.

Transcript Structure Effects on Transcript Extension. Before addressing this question, it is important to briefly review the available information on the structures of template·primer duplexes in cocrystals with polymerases and the effects of sequence, ribose structure, and solution conditions on duplex conformation. Crystal structures of polymerase·primer·template complexes have revealed that the duplexes in the active sites form structures that, while varying in details, are recognizably of the A- or B-form families (*39*, *40*). RNA·DNA hybrids favor A-form conformations (*1*, *14*−*16*, *40*, *41*). It has therefore been suggested that the RNA·DNA hybrid in the RNAP active site assumes an A-like conformation (*43*−*45*), and some limited direct evidence for this has been obtained (*46*). DNA·DNA duplexes favor B-form structures (*1*). On the basis of crystal structures it has been suggested that a single rNMP in a ∼10 base DNA·RNA chimeric duplex is sufficient to stabilize an A-form structure, but such studies all involve crystals grown from solutions containing high concentrations of alcohol or 2-methyl-2,4-pentanediol (MPD), conditions that are expected to destabilize B-form helices (*47*−*49*). Indeed, crystals of C·G-rich A-form DNA can be grown from such solutions (*16*). Structural studies of synthetic Okazaki fragments [such as r(ccca)d(AATGA)·d(TCATTTGGG)] in solution are more relevant to the case of an RNAP extending a mixed rNMP/dNMP transcript on a DNA template since they involve hybrids of pure DNA and chimeric RNA/DNA strands. Studies of these chimeras reveal B-type helices in the DNA·DNA segments, a 2−3 base pair junction region, and a hybrid helix form with varying degrees of A-like character in the RNA·DNA segment depending on its length and sequence (*42*, *48*−*50*). Such structures imply that while rNMPs may have a dominant effect on helix conformation, their presence in a hybrid is not overwhelming and RNA/DNA chimeras that are predominantly DNA are typically not A-form in solution. At the level of sequence it is known that dG·dC base pairs are more amenable to forming A-like structures than either dA·dT or dI·dC base pairs (*14*−*16*) and that alternating purine−pyrimidine sequences favor B-form to a greater degree than polypurine·polypyrimidine tracts (*17*). Uniquely amongst pure DNA sequences that have been studied, poly(dG)·poly(dC) is A-form in aqueous solution (*18*). A critical factor in the stabilization of the B-form structure is a highly ordered spine of hydrating water molecules in the minor groove (*15*, *16*, *51*). Disruption of this hydration spine by dehydration or by agents that disrupt water structure has been proposed to be the mechanism through which such environ-

mental conditions destabilize B-form DNA to favor A-form structures (*15*, *16*).

Against this background of information, our observations with T7 RNAP, while essentially correlative, strongly suggest that the barrier to extension of heavily dNMP-substituted transcripts is due to a change in the conformation of the transcript·template hybrid, most probably from an A-like structure to a B-like structure. Specifically we find the following: (1) dGTP is used efficiently as an initiating nucleotide, the presence of a 3′-dNMP in the transcript is not particularly limiting for transcript extension, and a moderate level of dNMP substitution in the transcript only modestly reduces extension efficiency. (2) The efficiency of transcript extension is strongly reduced when the 3−5 bases nearest the 3′-end of the transcript are predominantly deoxyribose. These observations appear inconsistent with any strong, direct discriminatory mechanism against utilization of a 3′-dNMP in the phosphoryl transfer reaction (in contrast to the discrimination observed for utilization of a dNTP vs an rNTP) and do not reveal any strong chemical specificity for a pure RNA transcript. They are consistent with a conformational effect caused by having predominantly deoxyribose content in the region of the transcript which is hybridized to the DNA template. (3) The activity of the Y639F enzyme is ≥600-fold lower for pure DNA than for pure RNA synthesis on a complex sequence template and DNA synthesis is not measurable on poly(dA)·poly(dT), but pure RNA or DNA synthesis activity is nearly identical on poly(dI)·poly(dC). (4) The activity of Y639F on a complex sequence template is much more sensitive to replacement of rGTP with dITP than with dGTP, though activity with rGTP vs rITP is identical. These observations are consistent with the idea of a requirement for an A-like structure in the hybrid and with the amenability with which dG·dC, dA·dT, and dI·dC base pairs form A-DNA. In particular, it appears difficult to account for the remarkably high activity of Y639F in pure DNA synthesis on the poly(dC) template without reference to the unique conformational properties of poly(dG)·poly(dC). (5) The addition of alcohols, and in particular methanol, greatly decreases the sensitivity of Y639F to replacement of multiple rNTPs with dNTPs in the transcription reaction, and this is due to enhanced extension of heavily dNMP-substituted transcripts rather than enhanced utilization of dNTPs. This would be consistent with the ability of such agents, through dehydration and water structure perturbing effects, to destabilize B-form helices, thereby allowing formation of an A-like structure that is a better substrate for extension.

The alcohol concentrations used in these experiments are, in fact, too low to induce a B → A transition in solution for most DNA sequences, and ethanol is actually more effective in inducing such a transition than methanol (*17*). It is therefore probable that the environment of the nucleic acid binding cleft in T7 RNAP also favors destabilization of a B-form conformation, either because the hydration shell of the nucleic acids is partially displaced upon association with the polymerase or because of specific chemical and shape complementarity between the binding cleft and an A-like conformation. Human immunodeficiency virus type I reverse transcriptase induces a DNA·DNA primer·template to assume an A-form conformation (*52*), and DNA polymerase β also causes a DNA·DNA primer·template to take on partial A-like character through interactions with protein side chains that disrupt the hydration spine of the minor groove (*40*).[2] For T7 RNAP, which normally works with an RNA·DNA hybrid, such features of the binding cleft may be inadequate to force an appropriate A-form structure with a DNA·DNA duplex, but together with other factors that also favor A-form conformations but are inadequate in isolation (moderate alcohol concentrations), may allow a DNA·DNA duplex to assume a conformation appropriate for efficient extension. We do not know why methanol is more effective than ethanol for this purpose, but it is possible that it is because methanol is less disruptive of polymerase activity than is ethanol (as assessed in a 4 rNTP reaction; Table 7) or because its smaller size allows it readier access to the transcript·template hybrid in the polymerase binding cleft.

It could be argued that alcohol addition exerts its effects in ways other than an effect on transcript·template conformation; for example, it could disrupt polymerase structure so that the enzyme becomes sloppier in its requirements for canonical substrate or transcript structure. Such structural effects on the enzyme are suggested by the depressing effect of alcohol addition on activity in 4 rNTP reactions. However, we have generally found that situations which lead to reductions in polymerase activity tend to *decrease* the relative utilization of noncanonical substrates even more (*5*). For example, nonconservative substitutions at position 639 that reduce activity typically enhance the rNTP/dNTP specificity of the polymerase relative to the conservative Y639F substitution (Table 1), and rGTP/dGTP specificity is also greater for the relatively inefficient dinucleotide synthesis reaction than it is during elongation (Tables 1 and 3). Such observations are consistent with the idea that the assumption of optimal catalytic geometry of the reactive groups in the active site involves a set of cooperative interactions between polymerase·template, polymerase·substrate, polymerase·primer/transcript, template·substrate, template·primer/transcript, and substrate·primer/transcript. Changes in any one of these interactions via a structural change in template, substrate, or the polymerase tend to affect kinetic parameters (*53*, *54*). Because of the highly nonlinear response of catalytic rates to changes in reactive group geometry, combinations of structural alterations (for example, the combination of a nonconservative active-site mutation and noncanonical substrate structure) may be expected to give rise to more than multiplicative effects on activity, as they are seen to do (Tables 1−3 or see ref *55*). We would therefore expect that any nonspecific disruptive effects of alcohols on enzyme structure that reduce activity in 4 rNTP reactions would tend to reduce activity in reactions with dNTPs to an even greater degree, which is obviously not what is found. In any case, it is clear that such effects on the polymerase, if they occur, do not lead to reduced specificity for canonical substrate structure at the level of nucleotide incorporation (Table 8).

It is possible that, in addition to this conformational mechanism, noncanonical transcript structure or the presence of a transcript binding site on the polymerase with some

[2] In fact, it is possible that most DNA-directed DNA polymerases will favor the formation of A-like structures in their active sites because the A-conformation is structurally more homogeneous with respect to sequence than is B-form DNA (*1*). Such structural homogeneity could be important since appropriate and restricted geometry of the base pairs and reactive groups is expected to be critical for the mechanisms of fidelity and catalysis.

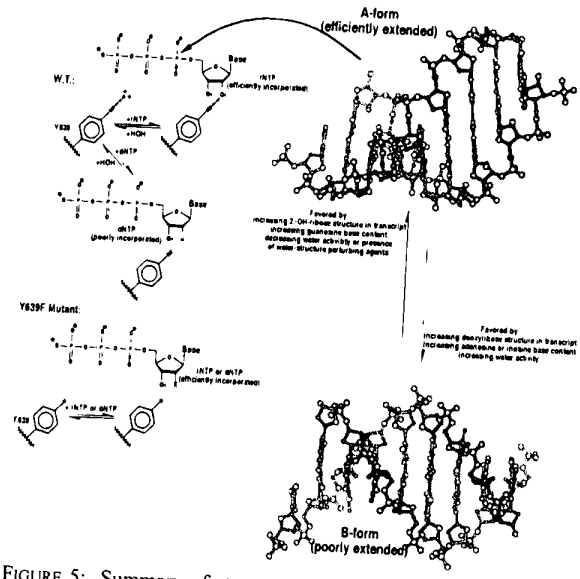


FIGURE 5: Summary of structural mechanisms underlying ribose specificity in RNA polymerization by T7 RNAP. The reaction involves two substrates: NTP and transcript. Specificity for the ribose 2′-group at the level of NTP utilization is mediated by the Y639 side chain as illustrated (*5*), via a mechanism analogous to that used by Y34 in tyrosyl-tRNA synthetase to discriminate tyrosine and phenylalanine substrates (*7*). Transcript structure effects on transcript extension are not confined to the 3′-NMP of the transcript and instead encompass *ca.* 5 bases (or more) at the transcript 3′-end. The correlation between the effects of alcohols, sequence, and ribose/deoxyribose structure on helix conformation (A- vs B-forms) and the effects of these factors on the efficiency of transcript extension imply a requirement for an A-like transcript·template hybrid for efficient extension. [A-form refers to all A-family type structures (A, A′, etc.); B-form refers to all B-family structures (B, C, D, etc.).]

chemical specificity for RNA contribute in other ways to the efficiency with which the transcript is extended. However, because we find that the synthesis of poly(dG) by Y639F is as efficient as poly(rG) synthesis, and because in 20% methanol the activity of Y639F in a 4 dNTP reaction is only ~10-fold less than in a 4 rNTP reaction, we believe that the primary determinant of extension efficiency with regard to the ribose/deoxyribose structure of the transcript is likely to be the helix conformation of the transcript·template hybrid.

The structural mechanisms that have been identified as contributing to substrate and product specificity in T7 RNAP are summarized in Figure 5. The emphasis that Figure 5 places on specific structural mechanisms is not meant to downplay the role played by the poorly processive initiation phase of transcription in limiting activity in reactions with noncanonical substrates. These have been previously discussed at length (*3*, *5*) and include both a kinetic component—a relatively modest elongation rate reduction during poorly processive initial transcription can drastically reduce activity by greatly reducing the fraction of elongation complexes that achieve promoter release—and a structural one—requirements for canonical substrate structure to achieve proper positioning of reactive groups may be more stringent during initial transcription because extended RNA·DNA and RNA·RNAP interactions that can contribute to positioning the 3′-NMP for catalysis are absent, leading, for example, to the greater sensitivity of the dinucleotide synthesis reaction to noncanonical ribose (Tables 6–8) or base (*5*, *56*) structure.

Given our understanding of the mechanisms of ribose discrimination, it should prove possible, for practical applications, to further develop this system for synthesis of nucleic acids with any desired level of 2′-substituted NMPs. For such purposes it is important that the Y639F mutation, Mn^{2+} substitution, and addition of methanol each affect the ribose specificity of the transcription reaction through largely nonoverlapping mechanisms so that these mechanisms can be combined to further enhance incorporation of 2′-substituted NMPs. For example, combination of ~15% methanol and the Y639F substitution allowed us, for the first time, to achieve significant levels of promoter-specific, unprimed DNA synthesis from a complex sequence template. This combination is not yet maximally efficient since methanol significantly reduces overall activity, but our observations suggest that nucleic acids with any desired level of 2′-modified ribose structure could be efficiently synthesized with the Y639F enzyme if solution conditions that destabilize B-form helices without depressing enzyme activity could be identified.

REFERENCES

1. Saenger, W. (1984) *Principles of Nucleic Acid Structure*, Springer-Verlag, New York.
2. Tabor, S., and Richardson, C. C. (1995) *Proc. Natl. Acad. Sci. U.S.A. 92*, 6339–6343.
3. Sousa, R., and Padilla, R. (1995) *EMBO J. 14*, 4609–4621.
4. Gao, G., Orlova, M., Georgiadis, M. M., Hendrickson, W. A., and Goff, S. P. (1997) *Proc. Natl. Acad. Sci. U.S.A. 94*, 407–411.
5. Huang, Y., Eckstein, F., Padilla, R., and Sousa, R. (1997) *Biochemistry 36*, 8231–8242.
6. Joyce, C. (1997) *Proc. Natl. Acad. Sci. U.S.A. 94*, 1619–1622.
7. Fersht, A. R., Shi, J.-P., Knill-Jones, J., Lowe, D. M., Wilkinson, A. J., Blow, D. M., Brick, P., Carter, P., Waye, M. M. Y., and Winter, G. (1985) *Nature 314*, 235–238.
8. Bonner, G., Patra, D., Lafer, E. M., and Sousa, R. (1992) *EMBO J. 11*, 3767–3775.
9. Tabor, S., and Richardson, C. C. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 1074–1078.
10. Kostyuk, D. A., Dragan, S. M., Lyakhov, D. L., Rechinsky, V. O., Tunitskaya, V. L., Chernov, B. K., and Kochetkov, S. N. (1995) *FEBS Lett. 369*, 165–168.
11. Martin, C. T., and Coleman, J. E. (1989) *Biochemistry 28*, 2760–2762.
12. Ivanov, V. I., Minchekova, L. E., Minyat, E. E., Frank-Kamenetskii, M. D., and Schyolkina, A. K. (1974) *J. Mol. Biol. 87*, 817–833.
13. Pohl, F. (1976) *Nature 260*, 365–366.
14. Leslie, A. G. W., Arnott, S., Chanrasekaran, R., and Ratliff, R. L. (1980) *J. Mol. Biol. 143*, 49–72.
15. Drew, H. R., and Dickerson, R. E. (1982) *Nature 295*, 294–299.
16. Drew, H. R., and Dickerson, R. E. (1981) *J. Mol. Biol. 151*, 535–556.
17. Arnott, S., Chandrasekaran, R., and Selsing, E. (1975) in *Structure and Conformation of Nucleic Acids and Protein–Nucleic Acid Interactions* (Sundaralingam, M., and Rao, S. T., Eds.) pp 577–596, University Park Press, Baltimore, MD.
18. Sarma, M. H., Gupta, G., and Sarma, R. H. (1986) *Biochemistry 25*, 3659–3665.
19. Ivanov, V. I., Minchekova, L. E., Schyolkina, A. K., and Poletayev, A. I. (1973) *Biopolymers 12*, 89–110.
20. Villemain, J., Guajardo, R., and Sousa, R. (1997) *J. Mol. Biol.* (submitted for publication).
21. Bonner, G., Lafer, E. M., and Sousa, R. (1994a) *J. Biol. Chem. 269*, 25120–25128.
22. Gardener, L. P., Mookhtiar, K. A., and Coleman, J. E. (1997) *Biochemistry 36*, 2908–2918.
23. Mookhtiar, K. A., Peluso, P. S., Muller, D. K., Dunn, J. J., and Coleman, J. E. (1991) *Biochemistry 30*, 6305–6313.

24. El-Deiry, W. S., Downey, K. M., and So, A. G. (1984) *Proc. Natl. Acad. Sci. U.S.A. 81*, 7378−7382.
25. Brandis, J., Edwards, S., and Jonhson, K. (1996) *Biochemistry 35*, 2189−2200.
26. Beckman, R. A., Mildvan, A. S., and Loeb, L. A. (1985) *Biochemistry 24*, 5810−5817.
27. Goodman, M. F., Keener, S., Guidotti, S., and Branscomb, E. W. (1983) *J. Biol. Chem. 258*, 3469−3475.
28. Pezo, V., and Wain-Hobson, S. (1997) *Gene 186*, 67−72.
29. Putrament, A., Baranowska, H., and Prazsmo, W. (1975) *Mol. Gen. Genet. 140*, 339−347.
30. Ripley, L. S. (1975) *Mol. Gen. Genet. 141*, 23−40.
31. Pinto, D., Sarocchi-Landousy, M. T., and Guschlbacher, W. (1979) *Nucleic Acids Res. 6*, 1041−1048.
32. Tabor, S., and Richardson, C. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 4076−4080.
33. Boiteux, S., and Laval, J. (1982) *Biochimie 64*, 637−641.
34. Ide, H., Kow, Y. W., and Wallace, S. S. (1985) *Nucleic Acids Res. 13*, 8035−8052.
35. Pelletier, H., Sawaya, M. R., Wolfle, W., Wilson, S. H., and Kraut, J. (1996) *Biochemistry 35*, 12762−12777.
36. Blasco, M. A., Lazaro, J. M., Bernard, A., Blanco, L., and Salas, M. (1992) *J. Biol. Chem. 267*, 19427−19434.
37. Inokuchi, Y., Kajitani, M., and Hirashima, A. (1994) *J. Biochem. (Tokyo) 116*, 1275−1280.
38. Scott, W. G., and Klug, A. (1996) *Trends Biochem. Sci. 21*, 220−224.
39. Irving, H., and Williams, R. (1953) *J. Am. Chem. Soc. 75*, 3193−3210.
40. Pelletier, H., Sawaya, M. R., Kumar, A., Wilson, S. H., and Kraut, J. (1994) *Science 264*, 1891−1903.
41. Arnott, S., Chandrasekaran, R., Millane, R. P., and Park, H. S. (1986) *J. Mol. Biol. 188*, 631−640.
42. Selsing, E., Wells, R. D., Early, T. A., and Kearns, D. R. (1978) *Nature 254*, 249−250.
43. Arnott, S., Fuller, W., Hodgson, A., and Prutton, I. (1968) *Nature 220*, 561−564.
44. Florentiev, V. L., and Ivanov, V. I. (1970) *Nature 228*, 519−522.
45. Hamilton, L. D. (1968) *Nature 218*, 633−637.
46. Beabealashvily, R. S., Ivanov, V. I., Minchenkova, L. E., and Savotchikina, L. P. (1972) *Biochim. Biophys. Acta 259*, 35−40.
47. Egli, M., Usman, N., Zhang, S., and Rich, A. (1991) *Proc. Natl. Acad. Sci. U.S.A. 89*, 534−538.
48. Egli, M., Usman, N., and Rich, A. (1993) *Biochemistry 32*, 3221−3237.
49. Wang, A. H. J., Fuji, S., van Boom, J. H., van der Marel, G. A., van Boeckel, S. A. A., and Rich, A. (1982) *Nature 299*, 601−604.
50. Salazar, M., Fedoroff, O. Y., and Reid, B. R. (1996) *Biochemistry 35*, 8126−8135.
51. Conner, B. N., Takano, T., Tanaka, S., Itakura, K., and Dickerson, R. E. (1982) *Nature 295*, 294−299.
52. Jacobo-Molina, A., Ding, J., Nanni, R. G., Clark, A. D., Jr., Lu, X., Tantillo, C., Williams, R. L., Kamer, G., Ferris, A. L., Clark, P., Hizi, A., Hughes, S. H., and Arnold, E. (1993) *Proc. Natl. Acad. Sci. U.S.A. 90*, 6320−6324.
53. Ricchetti, M., and Buc, H. (1993) *EMBO J. 12*, 387−396.
54. Tantillo, C., Ding, J., Jacobo-Molina, A., Nanni, R. G., Boyer, P. L., Hughes, S. H., Pauwels, R., Andries, K., Janssen, P. A. J., and Arnold, E. (1994) *Proc. Natl. Acad. Sci. U.S.A. 243*, 369−387.
55. Polesky, A. H., Steitz, T. A., Grindley, N. D. F., and Joyce, C. M. (1990) *J. Biol. Chem. 265*, 14579−14591.
56. Milligan, J. F., Groebe, D. R., Witherell, G. W., and Uhlenbeck, O. C. (1987) *Nucleic Acids Res. 15*, 8783−8798.

BI971609O